

3-2015

# QML Estimation of Dynamic Panel Data Models with Spatial Errors

Liangjun SU

Singapore Management University, [ljsu@smu.edu.sg](mailto:ljsu@smu.edu.sg)

Zhenlin YANG

Singapore Management University, [zlyang@smu.edu.sg](mailto:zlyang@smu.edu.sg)

**DOI:** <https://doi.org/10.1016/j.jeconom.2014.11.002>

Follow this and additional works at: [https://ink.library.smu.edu.sg/soe\\_research](https://ink.library.smu.edu.sg/soe_research)



Part of the [Econometrics Commons](#)

---

## Citation

SU, Liangjun and YANG, Zhenlin. QML Estimation of Dynamic Panel Data Models with Spatial Errors. (2015). *Journal of Econometrics*. 185, (1), 230-258. Research Collection School Of Economics.

**Available at:** [https://ink.library.smu.edu.sg/soe\\_research/1485](https://ink.library.smu.edu.sg/soe_research/1485)

This Journal Article is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [libIR@smu.edu.sg](mailto:libIR@smu.edu.sg).

# QML Estimation of Dynamic Panel Data Models with Spatial Errors\*

Liangjun Su and Zhenlin Yang<sup>†</sup>

School of Economics, Singapore Management University

November 6, 2014

## Abstract

We propose quasi maximum likelihood (QML) estimation of dynamic panel models with spatial errors when the cross-sectional dimension  $n$  is large and the time dimension  $T$  is fixed. We consider both the random effects and fixed effects models, and prove consistency and derive the limiting distributions of the QML estimators under different assumptions on the initial observations. We propose a residual-based bootstrap method for estimating the standard errors of the QML estimators. Monte Carlo simulation shows that both the QML estimators and the bootstrap standard errors perform well in finite samples under a correct assumption on initial observations, but may perform poorly when this assumption is not met.

**Key Words:** Bootstrap Standard Errors, Dynamic Panel, Fixed Effects, Initial Observations, Quasi Maximum Likelihood, Random Effects, Spatial Error Dependence.

**JEL Classification:** C10, C13, C21, C23, C15

## 1 Introduction

Recently, there has been a growing interest in the estimation of panel data models with cross-sectional or spatial dependence after Anselin (1988). See, among others, Elhorst (2003), Baltagi et al. (2003), Baltagi and Li (2004), Chen and Conley (2001), Pesaran (2004), Kapoor et al. (2007), Baltagi et al. (2007), Lee and Yu (2010a), Mutl and Pfaffermayr (2011), Parent and LeSage (2011), and Baltagi et

---

\*We thank Badi H. Baltagi, Harry Kelejian, Lung-fei Lee, Peter C. B. Phillips, Ingmar Prucha, and the seminar participants of Singapore Econometrics Study Group Meeting 2006, the 1st World Conference of the Spatial Econometric Association 2007, the 18th International Panel Data Conference 2012, and three anonymous referees and the co-editor Cheng Hsiao for their helpful comments. The early version of this paper was completed in 2007 when Su was with Peking University undertaking research supported by the grant NSFC (70501001). He thanks the School of Economics, Singapore Management University (SMU) for the hospitality during his two-month visit in 2006, and the Wharton-SMU research center for supporting his visit. He also acknowledges support from the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Zhenlin Yang gratefully acknowledges the research support from the Wharton-SMU research center.

<sup>†</sup>Corresponding Author: 90 Stamford Road, Singapore 178903. Phone: +65-6828-0852; Fax: +65-6828-0833. E-mail: [zlyang@smu.edu.sg](mailto:zlyang@smu.edu.sg).

al. (2013) for an overview on the static spatial panel data (SPD) models.<sup>1</sup> Adding a dynamic element into a SPD model further increases its flexibility, which has, since Anselin (2001), attracted the attention of many econometricians. The spatial dynamic panel data (SDPD) models can be broadly classified into two categories (Anselin, 2001, Anselin et al., 2008): one is that the dynamic and spatial effects both appear in the model in the forms of lags (in time and in space) of the response variable, and the other allows the dynamic effect in the same manner but builds the spatial effects into the disturbance term. The former has been studied by Yu et al. (2008), Yu and Lee (2010), Lee and Yu (2010b), and Elhorst (2010), and the latter by Elhorst (2005), Yang et al. (2006), Mutl (2006), and Su and Yang (2007). Lee and Yu (2010c) provide an excellent survey on the spatial panel data models (static and dynamic) and report some recent developments.

In this paper, we consider the latter type of SDPD model, in particular, the *dynamic panel data model with spatial error*. We focus on the more traditional panel data where the cross-sectional dimension  $n$  is allowed to grow but the time dimension  $T$  is held fixed (usually small), and follow the quasi-maximum likelihood (QML) approach for model estimation.<sup>2</sup> Elhorst (2005) studies the maximum likelihood estimation (MLE) of this model with fixed effects, but the asymptotic properties of the estimators are not given. Mutl (2006) investigates this model using the method of three-step generalized method of moments (GMM). Yang et al. (2006) consider a more general model where the response is subject to an unknown transformation and estimate the model by MLE. There are two well-known problems inherent from short panel and QML estimation, namely the *assumptions on the initial values* and the *incidental parameters*, and these problems remain for the SDPD model that we consider.<sup>3</sup> In the early version of this paper (Su and Yang, 2007), we derived the asymptotic properties of the QML estimators (QMLEs) of this model under both the random and fixed effects specifications with initial observations treated as either exogenous or endogenous, but methods for estimating the standard errors of the QMLEs were not provided. The main difficulty lies in the estimation of the variance-covariance (VC) matrix of the score function, where the traditional methods based on sample analogues, outer product of gradients, or analytical expressions fail due to the presence of error components in the original model and in the model for the initial observations. This difficulty is now overcome by a residual-based bootstrap method.

For over thirty years of spatial econometrics history, the asymptotic theory for the (Q)ML estimation of spatial models has been taken for granted until the influential paper by Lee (2004), which establishes systematically the desirable consistency and asymptotic normality results for the Gaussian QML estimates of a spatial autoregressive model. More recently, Yu et al. (2008) extend the work of Lee (2004) to spatial dynamic panel data models with fixed effects by allowing both  $T$  and  $n$  to be large. While our work is closely related to theirs, there are clear distinctions. First, unlike Yu et al. (2008) who consider only fixed effects model, we shall consider both random and fixed effects specifications of the individual

<sup>1</sup>For alternative approaches to model cross-sectional dependence, see Phillips and Sul (2003), Andrews (2005), Pesaran (2006), Bai (2009), Pesaran and Tosetti (2011), Su and Jin (2012), Moon and Weidner (2013), among others.

<sup>2</sup>A panel with large  $n$  and small  $T$ , called a short panel, remains the prevalent setting in the majority of empirical research involving many geographical regions or many economic agents, and evidence from the standard dynamic panel data models (Hsiao et al., 2002; Hsiao, 2003; Binder et al., 2005) and SDPD model with spatial lag (Elhorst, 2010) shows that QML estimators are more efficient than GMM estimators.

<sup>3</sup>See, for regular dynamic models, Balestra and Nerlove (1966), Nerlove (1971), Maddala (1971), Anderson and Hsiao (1981, 1982), Bhargava and Sargan (1983); Hsiao et al. 2002, Hsiao (2003), and Binder et al. (2005); and for spatial models, Su and Yang (2007), Elhorst (2010), and Parent and LeSage (2011).

effects. Second, we shall focus on the case of small  $T$ , and deal with the problems of *initial conditions* and *incidental parameters*. In contrast, neither problem arises under the large- $n$  and large- $T$  setting as considered in Yu et al. (2008). Third, spatial dependence is present only in the error term in our model whereas Yu et al. (2008) consider spatial lag model. It would be interesting to extend our work to the SDPD model with both spatial lag and spatial error.

To summarize, our paper provides a complete set of statistical inferences methodology to the small- $T$  SDPD model with spatial errors, accommodating different types of space-specific effects (random or fixed) and different ways that initial observations being generated (exogenously or endogenously). The proposed methods, including the bootstrap method for robust standard error estimation, are relatively easy to apply and thus greatly facilitates the empirical researchers. Yet, the main ideas are quite general and can be generalized to other types of SDPD models.

The rest of the paper is organized as follows. Section 2 introduces the basic model and discusses its extensions. Section 3 presents the QML estimation of the models with random or fixed effects, and exogenous or endogenous initial observations. The cases of endogenous initial observations are paid a specific attention where ‘predictive’ models are developed to ensure the information conveyed from the past are captured. Section 4 derives the asymptotic properties of the QMLEs. Section 5 introduces the bootstrap method for robust standard error estimation. Section 6 presents Monte Carlo results for the finite sample performance of the QMLEs and their estimated standard errors. Section 7 concludes the paper. All the proofs are relegated to the appendix.

*Notation.* For a positive integer  $k$ , let  $I_k$  denote a  $k \times k$  identity matrix,  $\iota_k$  a  $k \times 1$  vector of ones,  $0_k$  a  $k \times 1$  vector of zeros, and  $J_k = \iota_k \iota_k'$ , where  $'$  denotes transpose. Let  $A_1 \otimes A_2$  denote the Kronecker product of two matrices  $A_1$  and  $A_2$ . Let  $|\cdot|$ ,  $\|\cdot\|$ , and  $\text{tr}(\cdot)$  denote, respectively, the determinant, the Frobenius norm, and the trace of a matrix. We use  $\lambda_{\max}(A)$  and  $\lambda_{\min}(A)$  to denote the largest and smallest eigenvalues of a real symmetric matrix  $A$ .

## 2 Model Specification

We consider the SDPD model of the form

$$y_{it} = \rho y_{i,t-1} + x'_{it}\beta + z'_i\gamma + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (2.1)$$

where the scalar parameter  $\rho$  ( $|\rho| < 1$ ) characterizes the dynamic effect,  $x_{it}$  is a  $p \times 1$  vector of time-varying exogenous variables,  $z_i$  is a  $q \times 1$  vector of time-invariant exogenous variables that may include the constant term, dummy variables representing individuals’ gender, race, etc., and  $\beta$  and  $\gamma$  are the usual regression coefficients. The disturbance vector  $u_t = (u_{1t}, \dots, u_{nt})'$  is assumed to exhibit both non-observable individual effects and a spatially autocorrelated structure, i.e.,

$$u_t = \mu + \varepsilon_t, \quad (2.2)$$

$$\varepsilon_t = \lambda W_n \varepsilon_t + v_t, \quad (2.3)$$

where  $\mu = (\mu_1, \dots, \mu_n)'$ ,  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})'$ , and  $v_t = (v_{1t}, \dots, v_{nt})'$ , with  $\mu$  representing the unobservable individual or space-specific effects,  $\varepsilon_t$  representing the spatially correlated errors, and  $v_t$  representing the random innovations that are assumed to be independent and identically distributed (iid) with mean

zero and variance  $\sigma_v^2$ . The parameter  $\lambda$  is a spatial autoregressive coefficient and  $W_n$  is a known  $n \times n$  spatial weight matrix whose diagonal elements are zero.

Denoting  $y_t = (y_{1t}, \dots, y_{nt})'$ ,  $x_t = (x_{1t}, \dots, x_{nt})'$ , and  $z = (z_1, \dots, z_n)'$ , the model has the following reduced-form representation,

$$y_t = \rho y_{t-1} + x_t \beta + z \gamma + u_t, \text{ with } u_t = \mu + B_n^{-1} v_t, \quad t = 1, \dots, T, \quad (2.4)$$

where  $B_n = I_n - \lambda W_n$ . The following specifications are essential for the subsequent developments.

We focus on **short panels** where  $n \rightarrow \infty$  but  $T$  is fixed and typically small. Throughout the paper, the **initial observations** designated by  $y_0$  are considered to be available, which can be either **exogenous** or **endogenous**; the individual or space-specific effects  $\mu$  can be either ‘random’ or ‘fixed’, giving the so-called **random effects** and **fixed effects** models. To clarify, we adopt the view that the fundamental distinction between random effects and fixed effects models is not whether  $\mu$  is random or fixed, but rather whether  $\mu$  is uncorrelated or correlated with the observed regressors.

To give a unified presentation, we adopt a similar framework as Hsiao et al. (2002): (i) data collection starts from the 0th period; the process starts from the  $-m$ th period, i.e.,  $m$  periods before the start of data collection,  $m = 0, 1, \dots$ , and then evolves according to the model specified by (2.4); (ii) the starting position of the process  $y_{-m}$  is treated as exogenous; hence the exogenous variables  $(x_t, z)$  and the errors  $u_t$  start to have impact on the response from period  $-m + 1$  onwards; (iii) all exogenous quantities  $(y_{-m}, x_t, z)$  are considered as random and inferences proceed by conditioning on them, and (iv) variances of elements of  $y_{-m}$  are constant. Thus, when  $m = 0$ ,  $y_0 = y_{-m}$  is exogenous, when  $m \geq 1$ ,  $y_0$  becomes endogenous, and when  $m = \infty$ , the process has reached stationarity.

It is worth mentioning, in passing to model estimation, that although our model specified by (2.1)-(2.3) with random effects allows spatial dependence to be present only in the random disturbance term  $\varepsilon_t$  as in the static models considered by, e.g., Anselin (1988), Baltagi and Li (2004), and Baltagi et al. (2007), it can be easily extended to allow  $\mu$  to be spatially correlated in the same manner as  $\varepsilon_t$  (Kapoor et al., 2007), or to allow  $\mu$  to follow a different spatial process (Baltagi et al., 2013). See Section 3.1 for details. For ease of exposition we focus on the model specified by (2.1)-(2.3). When  $\mu$  represents fixed effects, as a referee kindly points out, these extensions do not make a difference in model estimation as fixed effects are wiped out by first differences.

### 3 The QML Estimators

In this section we develop quasi maximum likelihood estimates (QMLE) based on Gaussian likelihood for the SDPD model with random effects as well as the SDPD model with fixed effects. For the former, we start with the case of exogenous  $y_0$ , and then generalize it to give a unified treatment on the initial values. For the latter, a unified treatment is given directly.

#### 3.1 QMLEs for the random effects model

As indicated above, the main feature of the random effects SDPD model is that the state-specific effect  $\mu$  is assumed to be uncorrelated with the observed regressors. Furthermore, it is assumed that  $\mu$  contains iid elements of mean zero and variance  $\sigma_\mu^2$ , and is independent of  $v_t$ .

**Case I:  $y_0$  is exogenous ( $m = 0$ ).** In case when  $y_0$  is exogenous, it essentially contains no information with respect to the structural parameters in the system, and thus can be treated as fixed constants. In this case,  $x_0$  is not needed, and the estimation of the system makes use of  $T$  periods of data ( $t = 1, \dots, T$ ).

Conditional on the observed (exogenous)  $y_0$ , the distribution of  $y_1$  can be easily derived, and hence the Gaussian quasi-likelihood function based on the observations  $y_1, y_2, \dots, y_T$ . Define  $Y = (y'_1, \dots, y'_T)'$ ,  $Y_{-1} = (y'_0, \dots, y'_{T-1})'$ ,  $X = (x'_1, \dots, x'_T)'$ ,  $Z = \iota_T \otimes z$ , and  $v = (v'_1, \dots, v'_T)'$ . The SDPD model specified by (2.1)-(2.3) can be written in matrix form:

$$Y = \rho Y_{-1} + X\beta + Z\gamma + u, \quad \text{with } u = (\iota_T \otimes I_n)\mu + (I_T \otimes B^{-1})v. \quad (3.1)$$

Assuming  $\mu$  and  $v$  follow normal distributions leads to  $u \sim N(0, \sigma_v^2 \Omega)$ , where

$$\Omega \equiv \Omega(\lambda, \phi_\mu) = \phi_\mu(J_T \otimes I_n) + I_T \otimes (B'B)^{-1}, \quad (3.2)$$

$\phi_\mu = \sigma_\mu^2 / \sigma_v^2$ ,  $J_T = \iota_T \iota'_T$ , and  $B = B_n = I_n - \lambda W_n$ . Note that the dependence of  $B$  on  $n$  and  $\lambda$  is suppressed. The same notational convention is applied to other quantities such as  $Y$ ,  $X$ ,  $\Omega$ , etc., unless confusion arises.

The distribution of  $u$  leads to the distribution of  $Y - \rho Y_{-1}$ , and hence the distribution of  $Y$  as the Jacobian of the transformation is one. Let  $\theta = (\beta', \gamma', \rho)'$ ,  $\delta = (\lambda, \phi_\mu)'$ , and  $\psi = (\theta', \sigma_v^2, \delta')'$ . Denoting  $u(\theta) = Y - \rho Y_{-1} - X\beta - Z\gamma$ , the quasi-log-likelihood function of  $\psi$  is

$$\mathcal{L}^r(\psi) = -\frac{nT}{2} \log(2\pi) - \frac{nT}{2} \log(\sigma_v^2) - \frac{1}{2} \log |\Omega| - \frac{1}{2\sigma_v^2} u(\theta)' \Omega^{-1} u(\theta). \quad (3.3)$$

If the errors  $\{\mu_i\}$  and  $\{v_{it}\}$  are normally distributed, maximizing (3.3) gives the maximum likelihood estimator (MLE) of  $\psi$ . If they are not, but iid with mean zero, constant variances and, more importantly, finite fourth moments, maximizing (3.3) gives the QMLE of  $\psi$ . See Sections 4.1 and 4.2 for detailed regularity conditions. Given  $\delta$ , (3.3) is partially maximized at the concentrated QMLEs of  $\theta$  and  $\sigma_v^2$ ,

$$\hat{\theta}(\delta) = (\tilde{X}' \Omega^{-1} \tilde{X})^{-1} \tilde{X}' \Omega^{-1} Y \quad \text{and} \quad \hat{\sigma}_v^2(\delta) = \frac{1}{nT} \tilde{u}(\delta)' \Omega^{-1} \tilde{u}(\delta), \quad (3.4)$$

respectively, where  $\tilde{X} = (X, Z, Y_{-1})$  and  $\tilde{u}(\delta) = Y - \tilde{X} \hat{\theta}(\delta)$ . Substituting  $\hat{\theta}(\delta)$  and  $\hat{\sigma}_v^2(\delta)$  given in (3.4) back into (3.3) for  $\theta$  and  $\sigma_v^2$ , we obtain the concentrated quasi-log-likelihood function of  $\delta$ :

$$\mathcal{L}_c^r(\delta) = -\frac{nT}{2} [\log(2\pi) + 1] - \frac{nT}{2} \log[\hat{\sigma}_v^2(\delta)] - \frac{1}{2} \log |\Omega|. \quad (3.5)$$

The QMLE  $\hat{\delta} = (\hat{\lambda}, \hat{\phi}_\mu)'$  of  $\delta$  maximizes  $\mathcal{L}_c^r(\delta)$  given in (3.5). The QMLEs of  $\theta$  and  $\sigma_v^2$  are given by  $\hat{\theta} \equiv \hat{\theta}(\hat{\delta})$  and  $\hat{\sigma}_v^2 \equiv \hat{\sigma}_v^2(\hat{\delta})$ , respectively. Further, the QMLE of  $\sigma_\mu^2$  is given by  $\hat{\sigma}_\mu^2 = \hat{\phi}_\mu \hat{\sigma}_v^2$ .<sup>4</sup> Let  $\hat{\psi} = (\hat{\theta}', \hat{\sigma}_v^2, \hat{\delta}')'$ .

The QML estimation of the random effects SDPD model is seen to be very simple under exogenous  $y_0$ . The numerical maximization involves only two parameters, namely, the spatial parameter  $\lambda$  and the variance ratio  $\phi_\mu$ . The dynamic parameter  $\rho$  is estimated in the same way as the usual regression coefficients and its QMLE has an explicit expression given  $\lambda$  and  $\phi_\mu$ .

<sup>4</sup>As discussed at the end of Section 2, our results can easily be extended to allow  $\mu$  to be spatially correlated. For example, for Kapoor et al. (2007) model where  $u_t = \rho W u_t + \varepsilon_t$  and  $\varepsilon_t = \mu + v_t$ , all results go through with  $\Omega = \phi_\mu(J_T \otimes (B'B)^{-1}) + I_T \otimes (B'B)^{-1}$ ; for Baltagi et al. (2013) model where  $u_t = u_1 + u_{2t}$ ,  $u_1 = \rho_1 W_1 + \mu$ , and  $u_{2t} = \rho_2 W_2 u_{2t} + v_t$ , one simply replaces  $\Omega$  above by  $\Omega = \phi_\mu(J_T \otimes (B'_1 B_1)^{-1}) + I_T \otimes (B'_2 B_2)^{-1}$  where  $B_1 = I_n - \rho_1 W_1$  and  $B_2 = I_n - \rho_2 W_2$ .

**Case II:  $y_0$  is endogenous ( $m \geq 1$ ).** The log-likelihood function (3.3) is derived under the assumption that the initial observation  $y_0$  is exogenously given. If this assumption is not satisfied, maximizing (3.3) generally produces biased or inconsistent estimators (see Bhargava and Sargan, 1983, and Section 4.2 of this paper for details). On the other hand, if the initial observation  $y_0$  is taken as endogenous in the sense that it is generated from the process specified by (2.4), which starts  $m$  periods before the 0th period, then  $y_0$  contains useful information about the model parameters and hence should be utilized in the model estimation. In this case,  $x_0$  is needed, and the estimation makes use of  $T + 1$  periods of data. We now present a unified set-up for a general  $m$  and then argue (see Remark II below) that by letting  $m = 0$  it reduces to the case of exogenous  $y_0$ . By successive backward substitutions using (2.4), we have

$$y_0 = \rho^m y_{-m} + \sum_{j=0}^{m-1} \rho^j x_{-j} \beta + z \gamma \frac{1 - \rho^m}{1 - \rho} + \mu \frac{1 - \rho^m}{1 - \rho} + \sum_{j=0}^{m-1} \rho^j B^{-1} v_{-j}. \quad (3.6)$$

Letting  $\eta_0$  and  $\zeta_0$  be, respectively, the exogenous and endogenous components of  $y_0$ , we have

$$\eta_0 = \rho^m y_{-m} + \sum_{j=0}^{m-1} \rho^j x_{-j} \beta + z \gamma \frac{1 - \rho^m}{1 - \rho} = \eta_m + x_0 \beta + z_m(\rho) \gamma, \quad (3.7)$$

where  $\eta_m = \rho^m y_{-m} + \sum_{j=1}^{m-1} \rho^j x_{-j} \beta$  and  $z_m(\rho) = z \frac{1 - \rho^m}{1 - \rho}$ ; and

$$\zeta_0 = \mu \frac{1 - \rho^m}{1 - \rho} + \sum_{j=0}^{m-1} \rho^j B^{-1} v_{-j}, \quad (3.8)$$

where  $E(\zeta_0) = 0$  and  $\text{Var}(\zeta_0) = \sigma_\mu^2 \left( \frac{1 - \rho^m}{1 - \rho} \right)^2 I_n + \sigma_v^2 \frac{1 - \rho^{2m}}{1 - \rho^2} (B' B)^{-1}$ . Clearly, both the mean and variance of  $y_0$  are functions of the model parameters and hence  $y_0$  is informative to model estimation. Treating  $y_0$  as exogenous will lose such information and causes bias or inconsistency in model estimation.

However, both  $\{x_{-j}, j = 1, \dots, m-1\}$  for  $m \geq 2$  and  $y_{-m}$  for  $m \geq 1$  in  $\eta_m$  are unobserved, rendering that (3.7) cannot be used as a model for  $\eta_0$ . Some approximations are necessary. In this paper, we follow Bhargava and Sargan (1983) (see also Hsiao, 2003, p.76) and propose a model for the initial observations based on the following fundamental assumptions. Let  $\mathbf{x} \equiv (x_0, x_1, \dots, x_T)$ .

**Assumption R0:** (i) Conditional on the observables  $\mathbf{x}$  and  $z$ , the optimal predictors for  $x_{-j}, j \geq 1$ , are  $\mathbf{x}$  and the optimal predictors for  $E(y_{-m}), m \geq 1$ , are  $\mathbf{x}$  and  $z$ ; and (ii) The error resulted from predicting  $\eta_m$  using  $\mathbf{x}$  and  $z$  is  $\zeta$  such that  $\zeta \sim (0, \sigma_\zeta^2 I_n)$  and is independent of  $u$ ,  $\mathbf{x}$  and  $z$ .<sup>5</sup>

These assumptions lead immediately to the following model for  $\eta_m$ :

$$\eta_m = \iota_n \pi_1 + \mathbf{x} \pi_2 + z \pi_3 + \zeta \equiv \tilde{\mathbf{x}} \pi + \zeta, \quad (3.9)$$

where  $\tilde{\mathbf{x}} = (\iota_n, \mathbf{x}, z)$  and  $\pi = (\pi_1, \pi_2', \pi_3')'$ . Clearly, the variability of  $\zeta$  comes from two sources: the variability of  $y_{-m}$  and the variability of the prediction error from predicting  $E(y_{-m})$  and  $\sum_{j=1}^{m-1} \rho^j x_{-j} \beta$  by  $\mathbf{x}$  and  $z$ . Hence, we have the following model for  $y_0$  based on (3.6)-(3.9):

$$y_0 = \tilde{\mathbf{x}} \pi + x_0 \beta + z_m(\rho) \gamma + u_0, \quad u_0 = \zeta + \zeta_0. \quad (3.10)$$

<sup>5</sup>As a referee thoughtfully points out, it is possible to allow for additional spatial structure to characterize the initial observations. But this will surely complicate the asymptotic analysis and will add in more parameters to be estimated; we leave it for future work. Similarly remark holds for Assumption F0 in Section 3.2.

The ‘initial’ error vector  $u_0$  is seen to contain three components:  $\zeta$ ,  $\mu \frac{1-\rho^m}{1-\rho}$ , and  $\sum_{j=0}^{m-1} \rho^j B^{-1} v_{-j}$ , being, respectively, the prediction error from predicting the unobservables, the cumulative random effects up to the 0th period, and the ‘cumulative’ spatial effects and random shocks up to the 0th period. The term  $z_m(\rho)\gamma = z \frac{1-\rho^m}{1-\rho} \gamma$  represents the cumulative impact of the time-invariant variables  $z$  up to period 0 and needs not be predicted. However, the predictors for  $\eta_m$  still include  $z$ , indicating that (i) the mean of  $y_{-m}$  is allowed to be linearly related to  $z$  and (ii)  $\rho^m$  may not be small such that the effect of  $y_{-m}$  on  $\eta_m$  is not negligible. If  $\rho^m$  is small which occurs when either  $m$  is large or  $\rho$  is small, the impact of  $y_{-m}$  to  $\eta_m$  can be ignored, and the term  $z\pi_3$  involved in (3.10) should be omitted. Some details about the cases with small  $\rho^m$  are given latter. For the cases where  $\rho^m$  is not negligible, one can easily show that, under strict exogeneity of  $\mathbf{x}$  and  $z$ ,  $E(u_0) = 0$ ,

$$E(u_0 u_0') = \sigma_\zeta^2 I_n + \sigma_\mu^2 a_m^2 I_n + \sigma_v^2 b_m (B' B)^{-1}, \text{ and } E(u_0 u') = \sigma_\mu^2 a_m (\iota_T' \otimes I_n),$$

where  $a_m \equiv a_m(\rho) = \frac{1-\rho^m}{1-\rho}$  and  $b_m \equiv b_m(\rho) = \frac{1-\rho^{2m}}{1-\rho^2}$ . Let  $u^* = (u_0', u')'$ . Under the normality assumption for the original error components  $\mu$  and  $v$ , and the ‘new’ prediction error  $\zeta$ , we have  $u^* \sim N(0, \sigma_v^2 \Omega^*)$ , where  $\Omega^*$  is  $n(T+1) \times n(T+1)$  and has the form:

$$\Omega^* \equiv \Omega^*(\rho, \lambda, \phi_\mu, \phi_\zeta) = \begin{pmatrix} \phi_\zeta I_n + \phi_\mu a_m^2 I_n + b_m (B' B)^{-1} & \phi_\mu a_m (\iota_T' \otimes I_n) \\ \phi_\mu a_m (\iota_T \otimes I_n) & \Omega \end{pmatrix}, \quad (3.11)$$

$\phi_\zeta = \sigma_\zeta^2 / \sigma_v^2$ , and  $\Omega$  is given by (3.2). This leads to the joint distribution of  $(y_0', (Y - \rho Y_{-1})')'$ , and hence the joint distribution of  $(y_0', Y')'$  or the likelihood function. Again, the arguments of  $\Omega^*$  are frequently suppressed should no confusion arise.

Now let  $\theta = (\beta', \gamma', \pi')'$ ,  $\delta = (\rho, \lambda, \phi_\mu, \phi_\zeta)'$ , and  $\psi = (\theta', \sigma_v^2, \delta')'$ . Based on (2.4) and (3.10), the Gaussian quasi-log-likelihood function of  $\psi$  has the form:

$$\mathcal{L}^{rr}(\psi) = -\frac{n(T+1)}{2} \log(2\pi) - \frac{n(T+1)}{2} \log(\sigma_v^2) - \frac{1}{2} \log |\Omega^*| - \frac{1}{2\sigma_v^2} u^*(\theta, \rho)' \Omega^{*-1} u^*(\theta, \rho), \quad (3.12)$$

where  $u^*(\theta, \rho) = \{(y_0 - x_0\beta - z_m(\rho)\gamma - \tilde{\mathbf{x}}\pi)', (Y - \rho Y_{-1} - X\beta - Z\gamma)'\}' \equiv Y^* - X^*\theta$ ,

$$Y^* = Y^*(\rho) = \begin{pmatrix} y_0 \\ Y - \rho Y_{-1} \end{pmatrix} \quad \text{and} \quad X^* = X^*(\rho) = \begin{pmatrix} x_0 & z_m(\rho) & \tilde{\mathbf{x}} \\ X & Z & 0_{nT \times k} \end{pmatrix}.$$

Maximizing (3.12) gives MLE of  $\psi$  if the error components are truly Gaussian and the QMLE otherwise. Similar to **Case I**, we work with the concentrated quasi-log-likelihood by concentrating out the parameters  $\theta$  and  $\sigma_v^2$ . The constrained QMLEs of  $\theta$  and  $\sigma_v^2$ , given  $\delta$ , are

$$\hat{\theta}(\delta) = (X^{*'} \Omega^{*-1} X^*)^{-1} X^{*'} \Omega^{*-1} Y^* \quad \text{and} \quad \hat{\sigma}_v^2(\delta) = \frac{1}{n(T+1)} \tilde{u}^*(\delta)' \Omega^{*-1} \tilde{u}^*(\delta), \quad (3.13)$$

where  $\tilde{u}^*(\delta) = u^*(\hat{\theta}(\delta), \rho) = Y^* - X^* \hat{\theta}(\delta)$ , and  $\hat{\theta}(\delta) = (\hat{\beta}(\delta)', \hat{\gamma}(\delta)', \hat{\pi}(\delta)')'$ . Substituting  $\hat{\theta}(\delta)$  and  $\hat{\sigma}_v^2(\delta)$  back into (3.12) for  $\theta$  and  $\sigma_v^2$ , we obtain the concentrated quasi-log-likelihood function of  $\delta$ :

$$\mathcal{L}_c^{rr}(\delta) = -\frac{n(T+1)}{2} [\log(2\pi) + 1] - \frac{n(T+1)}{2} \log \hat{\sigma}_v^2(\delta) - \frac{1}{2} \log |\Omega^*|. \quad (3.14)$$

Maximizing  $\mathcal{L}_c^{rr}(\delta)$  given in (3.14) gives the QMLE of  $\delta$ , denoted by  $\hat{\delta} = (\hat{\rho}, \hat{\lambda}, \hat{\phi}_\mu, \hat{\phi}_\zeta)'$ . The QMLEs of  $\theta$  and  $\sigma_v^2$  are thus given by  $\hat{\theta} \equiv \hat{\theta}(\hat{\delta})$  and  $\hat{\sigma}_v^2 \equiv \hat{\sigma}_v^2(\hat{\delta})$ , respectively, and these of  $\sigma_\mu^2$  and  $\sigma_\zeta^2$  are given by



$\hat{\sigma}_\mu^2 = \hat{\phi}_\mu \hat{\sigma}_v^2$  and  $\hat{\sigma}_\zeta^2 = \hat{\phi}_\zeta \hat{\sigma}_v^2$ , respectively.<sup>6</sup> Let  $\hat{\psi} = (\hat{\theta}', \sigma_v^2, \hat{\delta}')'$ .

**Remark I:** To utilize the information contained in the  $n$  initial observations  $y_0$ , we have introduced  $k = p(T+1) + q + 1$  additional parameters  $(\pi, \sigma_\zeta^2)$  in the model (3.9). Besides the bias issue, efficiency gain by utilizing additional  $n$  observations is reflected by  $n - k$ . Apparently, the condition  $n > k$  has to be satisfied in order for  $\pi$  and  $\sigma_\zeta^2$  to be identified. If both  $T$  and  $p$  are not so small ( $T = 9$  and  $p = 10$ , say), one may consider replacing the regressors  $\mathbf{x}$  in (3.9) by the most relevant ones (to the past),  $x_0$  and  $x_1$ , say, or simply by  $\bar{x} = (T+1)^{-1} \sum_{t=0}^T x_t$ . In this case  $k = 2p + q + 1$ , and  $p + q + 1$ , respectively. See Elhorst (2010) for similar remarks for an SDPD model with a spatial lag.

**Remark II:** When  $y_0$  is exogenous, model (3.10) becomes  $y_0 = \tilde{\mathbf{x}}\pi + u_0$ , where  $u_0 \sim (0, \sigma_0^2 I_n)$  and is independent of  $u$ . In this case, we have  $\Omega^* = \text{diag}(\sigma_0^2 I_n, \Omega)$ . Model estimation may proceed by letting  $m = 0$  in (3.14), and the results are almost identical to those from maximizing (3.5). A special case of this is the one considered in Hsiao (2003, p.76, Case IIa) where  $y'_{i0}$ s are simply assumed to be iid independent of  $\mu_i$ . If  $y'_{i0}$ s are allowed to be correlated with  $\mu'_i$  (Case IIb, Hsiao, 2003, p.76), the model becomes a special case of endogenous  $y_0$  as considered above.

**Remark III:** In general,  $m$  is unknown. In dealing with a dynamic panel model with fixed effects but without spatial dependence, Hsiao et al. (2002) recommend treating  $m$  or a function of it as a free parameter, which is estimated jointly with the other model parameters. However, we note that their approach requires  $\rho \neq 0$ , as when  $\rho = 0$ ,  $m$  disappears from the model and hence cannot be identified. Elhorst (2005) recommends that an appropriate value of  $m$  should be chosen in advance. We concur with his view for two reasons: (i) an empirical study often tells roughly what the  $m$  value is (see, e.g., the application considered by Elhorst), and (ii) the estimation is often not sensitive to the choice of  $m$  unless it is very small ( $m \leq 2$ ), and  $|\rho|$  is close to 1, as evidenced by the Monte Carlo results given in Section 6.

While the results given above are under a rather general set-up, some special cases deserve detailed discussions, which are (a)  $m = 1$ , (b)  $m = \infty$ , and (c)  $\rho = 0$ .

**(a)  $m=1$ .** When the process starts just one period before the start of data collection, the model (3.10) becomes  $y_0 = \rho y_{-1} + x_0 \beta + z \gamma + \mu + B^{-1} v_0$ ,  $z_m(\rho) = z$ , and

$$\Omega^* = \begin{pmatrix} (\phi_\zeta + \phi_\mu) I_n + (B' B)^{-1}, & \phi_\mu (\iota'_T \otimes I_n) \\ \phi_\mu (\iota'_T \otimes I_n), & \Omega \end{pmatrix}.$$

In this case,  $\rho$  becomes a linear parameter again and the estimation can be simplified by putting  $\rho$  together with  $\beta$ ,  $\gamma$  and  $\pi$  which can be concentrated out from the likelihood function. Now, denoting the response vector and the regressor matrix by:

$$\tilde{Y} = \begin{pmatrix} y_0 \\ Y \end{pmatrix} \quad \text{and} \quad \tilde{X} = \begin{pmatrix} x_0 & z & 0_{n \times 1} & \tilde{\mathbf{x}} \\ X & Z & Y_{-1} & 0_{nT \times k} \end{pmatrix},$$

the estimation proceeds with  $\theta = (\beta', \gamma', \rho, \pi)'$  and  $\delta = (\lambda, \phi_\mu, \phi_\zeta)'$ .

**(b)  $m=\infty$ .** When the process has reached stationarity ( $m \rightarrow \infty$  and  $|\rho| < 1$ ), the model for the initial observations becomes  $y_0 = \sum_{j=0}^{\infty} \rho^j x_{-j} \beta + \frac{z \gamma}{1-\rho} + \frac{\mu}{1-\rho} + \sum_{j=0}^{\infty} \rho^j B^{-1} v_{-j}$ . As  $\eta_\infty = \sum_{j=0}^{\infty} \rho^j x_{-j} \beta$  involves

<sup>6</sup>Unlike the case of exogenous  $y_0$ , the dynamic parameter  $\rho$  now becomes a nonlinear parameter that has to be estimated, together with  $\lambda$ ,  $\phi_\mu$  and  $\phi_\zeta$ , through a nonlinear optimization process. Similar to the case of exogenous  $y_0$ , our model and estimation can easily be extended to allow  $\mu$  to be spatially correlated as in Kapoor et al. (2007), or Baltagi et al. (2013).

only the time-varying regressors, its optimal predictors should be  $(\iota_n, \mathbf{x})$ . The estimation proceeds by letting  $z_m(\rho) = z_\infty(\rho) = \frac{z}{1-\rho}$ ,  $a_m = a_\infty = \frac{1}{1-\rho}$ ,  $b_m = b_\infty = \frac{1}{1-\rho^2}$ ,  $\tilde{\mathbf{x}} = (\iota, \mathbf{x})$ , and  $\pi = (\pi_1, \pi_2)'$ .

(c)  $\rho = 0$ . When the true value of the dynamic parameter is zero, the model becomes static with  $y_t = x_t\beta + z\gamma + \mu + B^{-1}v_t$ ,  $t = 0, 1, \dots, T$ . At this point, the true values for all the added parameters,  $\pi$  and  $\sigma_\zeta$ , are automatically zero.

### 3.2 QMLEs for the fixed effects model

In this section, we consider the QML estimation of the SDPD model with fixed effects, i.e., the vector of unobserved individual-specific effects  $\mu$  in model (2.4) is allowed to correlate with the time-varying regressors  $x_t$ . Due to this unknown correlation,  $\mu$  acts as if they are  $n$  free parameters, and with  $T$  fixed the model cannot be consistently estimated due to the incident parameter problem. Following the standard practice, we eliminate  $\mu$  by first-differencing (2.4) to give

$$\Delta y_t = \rho \Delta y_{t-1} + \Delta x_t \beta + \Delta u_t, \quad \Delta u_t = B^{-1} \Delta v_t, \quad t = 2, 3, \dots, T. \quad (3.15)$$

Clearly, (3.15) is not defined for  $t = 1$  as  $\Delta y_1$  depends on  $\Delta y_0$  and the latter is not observed. Thus, even if  $y_0$  (hence  $\Delta y_0$ ) is exogenous, one cannot formulate the likelihood function by conditioning on  $\Delta y_0$  as in the early case. To obtain the joint distribution of  $\Delta y_1, \Delta y_2, \dots, \Delta y_T$  or the transformed likelihood function for the remaining parameters based on (3.15), a proper approximation for  $\Delta y_1$  needs to be made so that its marginal distribution can be obtained, whether  $y_0$  is exogenous or endogenous. We present a unified treatment for the fixed effects model where the initial observations  $y_0$  can be exogenous ( $m = 0$ ) as well as endogenous ( $m \geq 1$ ).

Under the general specifications given at the end of Section 2, continuous backward substitutions to the previous  $m(\geq 1)$  periods leads to

$$\Delta y_1 = \rho^m \Delta y_{-m+1} + \sum_{j=0}^{m-1} \rho^j \Delta x_{1-j} \beta + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j}. \quad (3.16)$$

Note that (i)  $\Delta y_{-m+1}$  represents the changes after the process has made its first move, called the *initial endowment*; (ii) while the starting position  $y_{-m}$  is assumed exogenous, the initial endowment  $\Delta y_{-m+1}$  is endogenous, and (iii) when  $m = 0$ ,  $\Delta y_{-m+1} = \Delta y_1$ , i.e., the initial endowment becomes the *observed initial difference*. The effect of the initial endowment decays as  $m$  increases. However, when  $m$  is small, their effect can be significant, and hence a proper approximation to it is important. In general, write  $\Delta y_1 = \Delta \eta_1 + \Delta \zeta_1$ , where  $\Delta \eta_1$  and  $\Delta \zeta_1$ , the exogenous and endogenous components of  $\Delta y_1$ , are given as

$$\Delta \eta_1 = \rho^m E(\Delta y_{-m+1}) + \sum_{j=0}^{m-1} \rho^j \Delta x_{1-j} \beta \equiv \eta_m + \Delta x_1 \beta, \quad (3.17)$$

$$\Delta \zeta_1 = \rho^m [\Delta y_{-m+1} - E(\Delta y_{-m+1})] + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j}, \quad (3.18)$$

where  $\eta_m = \rho^m E(\Delta y_{-m+1}) + \sum_{j=1}^{m-1} \rho^j \Delta x_{1-j} \beta$ . Note that when  $m = 0$ , the summation terms in (3.17) and (3.18) should vanish, and as a result  $\Delta \eta_1 = E(\Delta y_1)$  and  $\Delta \zeta_1 = \Delta y_1 - E(\Delta y_1)$ .

Clearly, the observations  $\Delta x_{1-j}$ ,  $j = 1, \dots, m-1$ ,  $m \geq 2$ , are not available, and the structure of  $E(\Delta y_{-m+1})$ ,  $m \geq 1$ , is unknown. Hence  $\eta_m$  is completely unknown. Furthermore, as  $\eta_m$  is an

$n \times 1$  vector, it cannot be treated as a free parameter vector to be estimated; otherwise the incidental parameters problem will be confronted again.<sup>7</sup> Hsiao et al. (2002) remark that to get around this problem, the expected value of  $\Delta\eta_1$ , conditional on the observables, has to be a function of a finite number of parameters, and that such a condition can hold provided that  $\{x_{it}\}$  are trend-stationary (with a common deterministic linear trend) or first-difference stationary processes. Letting  $\Delta\mathbf{x} = (\Delta x_1, \dots, \Delta x_T)$ , we have the following fundamental assumptions.

**Assumption F0:** (i) The optimal predictors for  $\Delta x_{1-j}$ ,  $j = 1, 2, \dots$  and  $E(\Delta y_{-m+1})$ ,  $m = 0, 1, \dots$ , conditional on the observables, are  $\Delta\mathbf{x}$ ; (ii) Collectively, the errors from using  $\Delta\mathbf{x}$  to predict  $\eta_m$  is  $\epsilon \sim (0, \sigma_\epsilon^2 I_n)$ , and (iii)  $y_{-m} = E(y_{-m}) + e$ , where  $e \sim (0, \sigma_e^2 I_n)$  independent of  $\epsilon$ .

Assumption F0(i) and Assumption F0(ii) lead immediately to a ‘predictive’ model for  $\eta_m$ :

$$\eta_m = \pi_1 \iota_n + \Delta\mathbf{x} \pi_2 + \epsilon \equiv \tilde{\Delta}\mathbf{x} \pi + \epsilon, \quad m = 0, 1, \dots,$$

where  $\tilde{\Delta}\mathbf{x} = (\iota_n, \Delta\mathbf{x})$  and  $\pi = (\pi_1, \pi_2)'$ . Thus,  $\Delta\eta_1$  defined in (3.17) can be predicted by:  $\Delta\eta_1 = \tilde{\Delta}\mathbf{x} \pi + \Delta x_1 \beta + \epsilon$ . The original theoretical model (2.4) and Assumption F0(iii) lead to

$$\Delta y_{-m+1} - E(\Delta y_{-m+1}) = B^{-1} v_{-m+1} - (1 - \rho)e, \quad m = 0, 1, \dots,$$

which gives  $\Delta\zeta_1 = -\rho^m(1 - \rho)e + \rho^m B^{-1} v_{-m+1} + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j}$  when  $m \geq 1$ , and  $-(1 - \rho)e + B^{-1} v_1$  when  $m = 0$ . We thus have the following model for the observed initial difference,

$$\Delta y_1 = \tilde{\Delta}\mathbf{x} \pi + \Delta x_1 \beta + \epsilon + \Delta\zeta_1 \equiv \tilde{\Delta}\mathbf{x} \pi + \Delta x_1 \beta + \Delta\tilde{u}_1, \quad (3.19)$$

where  $\Delta\tilde{u}_1 = \epsilon + \Delta\zeta_1 = \epsilon - \rho^m(1 - \rho)e + \rho^m B^{-1} v_{-m+1} + \sum_{j=0}^{m-1} \rho^j B^{-1} \Delta v_{1-j}$ . Let  $\zeta = \epsilon - \rho^m(1 - \rho)e$ . By assumption, the elements of  $\zeta$  are iid with mean zero and variance  $\sigma_\zeta^2 = \sigma_\epsilon^2 + \sigma_e^2 \rho^{2m}(1 - \rho)^2$ .<sup>8</sup>

By construction, we can verify that under strict exogeneity of  $x_{it}$ , i.e.,  $E(\zeta_i | \Delta x_{i,1}, \dots, \Delta x_{i,T}) = 0$ , and independence between  $\zeta$  and  $\{\Delta v_{1-j}, j = 0, 1, \dots, m-1\}$ ,

$$E(\Delta\tilde{u}_1 \Delta\tilde{u}_1') = \sigma_\zeta^2 I_n + \sigma_v^2 c_m (B' B)^{-1} = \sigma_v^2 B^{-1} (\phi_\zeta B B' + c_m I_n) B'^{-1}, \quad \text{and} \quad (3.20)$$

$$E(\Delta\tilde{u}_1 \Delta u_t') = -\sigma_v^2 (B' B)^{-1} \quad \text{for } t = 2, \quad \text{and } 0 \text{ for } t = 3, 4, \dots, T, \quad (3.21)$$

where  $c_m \equiv c_m(\rho) = \frac{2}{1+\rho} - \frac{\rho^{2m}(1-\rho)}{1+\rho}$  and  $\phi_\zeta = \sigma_\zeta^2 / \sigma_v^2$ . Note that  $c_0 = 1$ ,  $c_\infty = \frac{2}{1+\rho}$  and  $c_m(0) = 2$ .

Letting  $\Delta u = (\Delta\tilde{u}_1', \Delta u_2', \dots, \Delta u_T')$ , we have  $\text{Var}(\Delta u) = \sigma_v^2 \Omega^\dagger$ , where

$$\Omega^\dagger \equiv \Omega^\dagger(\rho, \lambda, \phi_\zeta) = (I_T \otimes B^{-1}) H_E (I_T \otimes B'^{-1}), \quad (3.22)$$

$E = \phi_\zeta B B' + c_m I_n$ , and  $H_E$  is an  $nT \times nT$  matrix defined as

$$H_E = \begin{pmatrix} E & -I_n & 0 & \cdots & 0 & 0 & 0 \\ -I_n & 2I_n & -I_n & \cdots & 0 & 0 & 0 \\ 0 & -I_n & 2I_n & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2I_n & -I_n & 0 \\ 0 & 0 & 0 & \cdots & -I_n & 2I_n & -I_n \\ 0 & 0 & 0 & \cdots & 0 & -I_n & 2I_n \end{pmatrix}. \quad (3.23)$$

<sup>7</sup>An exception occurs when Model (2.4) does not contain time-varying variables as in Anderson and Hsiao (1981).

<sup>8</sup>Note that when  $m = 0$ ,  $\Delta\tilde{u}_1 = \epsilon - (1 - \rho)e + B^{-1} v_1$ . The approximation (3.19) is associated with Bhargava and Sargan’s (1983) approximation for the standard dynamic random effects model with endogenous initial observations. See Ridder and Wansbeek (1990) and Blundell and Smith (1991) for a similar approach.

The expression for  $\Omega^\dagger$  given in (3.22) greatly facilitates the calculation of the determinant and inverse of  $\Omega^\dagger$  as seen in the subsequent subsection. Derivations of score and Hessian matrix requires the derivatives of  $\Omega^\dagger$ , which can be made much easier based on the following alternative expression

$$\Omega^\dagger = \phi_\zeta(\ell_1 \otimes I_n) + h_{c_m} \otimes (B'B)^{-1}, \quad (3.24)$$

where  $\ell_1$  is a  $T \times T$  matrix with 1 in its top-left corner and zero elsewhere, and  $h_{c_m}$  is  $h_s$  defined at the end of Section 3.3 with  $s$  replaced by  $c_m$ .

In the following, we simply refer to the dimension of  $\pi$  to be  $k$ . Now let  $\theta = (\beta', \pi')'$ ,  $\delta = (\rho, \lambda, \phi_\zeta)'$ , and  $\psi = (\theta', \sigma_v^2, \delta')'$ . Note that  $\psi$  is a  $(p+k+4) \times 1$  vector of unknown parameters. Based on (3.15) and (3.19), the Gaussian quasi-log-likelihood of  $\psi$  has the form:

$$\mathcal{L}^f(\psi) = -\frac{nT}{2} \log(2\pi) - \frac{nT}{2} \log(\sigma_v^2) - \frac{1}{2} \log |\Omega^\dagger| - \frac{1}{2\sigma_v^2} \Delta u(\theta, \rho)' \Omega^{\dagger-1} \Delta u(\theta, \rho), \quad (3.25)$$

where  $\Delta u(\theta, \rho) = \Delta Y^\dagger(\rho) - \Delta X^\dagger \theta$ ,

$$\Delta Y^\dagger(\rho) = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 - \rho \Delta y_1 \\ \vdots \\ \Delta y_T - \rho \Delta y_{T-1} \end{pmatrix}, \quad \text{and} \quad \Delta X^\dagger = \begin{pmatrix} \Delta x_1 & \tilde{\Delta} \mathbf{x} \\ \Delta x_2 & 0_{n \times k} \\ \vdots & \vdots \\ \Delta x_T & 0_{n \times k} \end{pmatrix}.$$

Maximizing (3.25) gives the Gaussian MLE or QMLE of  $\psi$ . First, given  $\delta = (\rho, \lambda, \phi_\zeta)'$ , the constrained MLEs or QMLEs of  $\theta$  and  $\sigma_v^2$  are, respectively,

$$\hat{\theta}(\delta) = (\Delta X^\dagger' \Omega^{\dagger-1} \Delta X^\dagger)^{-1} \Delta X^\dagger' \Omega^{\dagger-1} \Delta Y^\dagger(\rho) \quad \text{and} \quad \hat{\sigma}_v^2(\delta) = \frac{1}{nT} \tilde{\Delta} u(\delta)' \Omega^{\dagger-1} \tilde{\Delta} u(\delta), \quad (3.26)$$

where  $\tilde{\Delta} u(\delta)$  equals  $\Delta u(\theta, \rho)$  with  $\theta$  being replaced by  $\hat{\theta}(\delta)$ . Substituting  $\hat{\theta}(\delta)$  and  $\hat{\sigma}_v^2(\delta)$  back into (3.25) for  $\theta$  and  $\sigma_v^2$ , we obtain the concentrated quasi-log-likelihood function of  $\delta$ :

$$\mathcal{L}_c^f(\delta) = -\frac{nT}{2} [\log(2\pi) + 1] - \frac{nT}{2} \log \hat{\sigma}_v^2(\delta) - \frac{1}{2} \log |\Omega^\dagger|. \quad (3.27)$$

The QMLE  $\hat{\delta} = (\hat{\rho}, \hat{\lambda}, \hat{\phi}_\zeta)'$  of  $\delta$  maximizes  $\mathcal{L}_c^f(\delta)$  given in (3.27). The QMLEs of  $\theta$  and  $\sigma_v^2$  are given by  $\hat{\theta} \equiv \hat{\theta}(\hat{\delta})$  and  $\hat{\sigma}_v^2 \equiv \hat{\sigma}_v^2(\hat{\delta})$ , respectively. Further, the QMLE of  $\sigma_\zeta^2$  are given by  $\hat{\sigma}_\zeta^2 = \hat{\phi}_\zeta \hat{\sigma}_v^2$ .<sup>9</sup> Let  $\hat{\psi} = (\hat{\theta}', \hat{\sigma}_v^2, \hat{\delta}')'$ .

**Remark IV:** We require that  $n > pT + 1$  for the identification of the parameters in (3.19). When this is too demanding, it can be addressed in the same manner as in the random effects model by choosing variables  $\Delta \tilde{\mathbf{x}}$  with a smaller dimension. For example, replacing  $\Delta \mathbf{x}$  in (3.19) by  $\overline{\Delta \mathbf{x}} = T^{-1} \sum_{t=1}^T \Delta x_t$  gives  $\Delta \tilde{\mathbf{x}} = (\iota_n, \overline{\Delta \mathbf{x}})$ , and dropping  $\Delta \mathbf{x}$  in (3.19) gives  $\tilde{\Delta} \mathbf{x} = \iota_n$ . In each case, the variance-covariance structure of  $\Delta u$  remains the same.

**Remark V:** Hsiao et al. (2002, p.110), in dealing with a dynamic panel data model without spatial effect, recommend treating  $c_m(\rho)$  as a free parameter to be estimated together with other model parameters. This essentially requires that  $\rho \neq 0$  and  $m$  be an unknown finite number. Note that  $c_m(0) = 2$

<sup>9</sup>Model (3.15) can be estimated by a simpler three-step IV-GMM type procedure suggested by Mutl (2006). When  $T$  is small the QMLE may be more efficient as it uses an extra period data, but the three-step procedure is free of initial conditions. Nevertheless, it should be interesting, as a future research, to conduct a formal comparison of the two models.

and  $c_\infty(\rho) = 2/(1 + \rho)$ , which become either a constant or a pure function of  $\rho$ . Our set-up allows  $\rho = 0$  or  $m = \infty$  so that a test for the existence of dynamics can be carried out or a stationary model can be fit. As in the case of the random effects model, we again treat  $m$  as known, chosen in advance based on the given data (see Remark III given in section 3.2).

### 3.3 Some computational notes

Maximization of  $\mathcal{L}_c^r(\delta)$ ,  $\mathcal{L}_c^{rr}(\delta)$  and  $\mathcal{L}_c^f(\delta)$  involves repeated evaluations of the inverse and determinants of the  $nT \times nT$  matrices  $\Omega$  and  $\Omega^\dagger$ , and the  $n(T+1) \times n(T+1)$  matrix  $\Omega^*$ . This can be a great burden when  $n$  or  $T$  or both are large. By Magnus (1982, p.242), the following identities can be used to simplify the calculation involving  $\Omega$  defined in (3.2):

$$|\Omega| = |(B'B)^{-1} + \phi_\mu T I_n| \cdot |B|^{-2(T-1)}, \quad (3.28)$$

$$\Omega^{-1} = T^{-1} J_T \otimes ((B'B)^{-1} + \phi_\mu T I_n)^{-1} + (I_T - T^{-1} J_T) \otimes (B'B). \quad (3.29)$$

The above formulae reduce the calculations of the inverse and determinant of an  $nT \times nT$  matrix to the calculations of those of several  $n \times n$  matrices, where the key element is the  $n \times n$  matrix  $B$ . By Griffith (1988), calculations of the determinants can be further simplified as:

$$|B| = \prod_{i=1}^n (1 - \lambda w_i), \text{ and } |(B'B)^{-1} + \phi_\mu T I_n| = \prod_{i=1}^n [(1 - \lambda w_i)^{-2} + \phi_\mu T], \quad (3.30)$$

where  $w_i$ 's are the eigenvalues of  $W$ . The above simplifications are also used in Yang et al. (2006).

For the determinant and inverse of  $\Omega^*$  defined in (3.11), let  $\omega_{11} = \phi_\zeta I_n + \phi_\mu a_m^2 I_n + b_m (B'B)^{-1}$ ,  $\omega_{21} = \omega'_{12} = \phi_\mu a_m (\iota_T \otimes I_n)$ , and  $D = \omega_{11} - \omega_{12} \Omega^{-1} \omega_{21}$ . We have by using the formulas for a partitioned matrix (e.g., Magnus and Neudecker, 2002, p.106),  $|\Omega^*| = |\Omega| \cdot |D|$ , and

$$\Omega^{*-1} = \begin{pmatrix} D^{-1} & -D^{-1} \omega_{12} \Omega^{-1} \\ -\Omega^{-1} \omega_{21} D^{-1} & \Omega^{-1} + \Omega^{-1} \omega_{21} D^{-1} \omega_{12} \Omega^{-1} \end{pmatrix}. \quad (3.31)$$

Thus, the calculations of the determinant and inverse of the  $n(T+1) \times n(T+1)$  matrix  $\Omega^*$  are reduced to the calculations of those of the  $n \times n$  matrix  $D$ , and those of  $\Omega$  given in (3.28) and (3.29).

For the determinant and inverse of  $\Omega^\dagger$  defined in (3.22), by the properties of matrix operation,

$$\begin{aligned} |\Omega^\dagger| &= |(I_T \otimes B^{-1})| \cdot |H_E| \cdot |(I_T \otimes B'^{-1})| = |B|^{-2T} |H_E|, \\ \Omega^{\dagger-1} &= (I_T \otimes B'^{-1})^{-1} H_E^{-1} (I_T \otimes B^{-1})^{-1} = (I_T \otimes B') H_E^{-1} (I_T \otimes B), \end{aligned}$$

where  $|H_E| = |TE - (T-1)I_n| = \prod_{i=1}^n [T\phi_\zeta(1 - \lambda w_i)^2 + Tc_m - T + 1]$  as in (3.30), and

$$H_E^{-1} = (1 - T)(h_0^{-1} \otimes E^{*-1}) + (h_1^{-1} - (1 - T)h_0^{-1}) \otimes (E^{*-1}E), \quad (3.32)$$

where  $E^* = TE - (T-1)I_n$ , and the  $T \times T$  matrices  $h_s, s = 0, 1$ , are

$$h_s = \begin{pmatrix} s & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix},$$

as in Hsiao et al. (2002, Appendix B), who also give  $|h_s| = 1 + T(s-1)$  and the expression for  $h_s^{-1}$ .

## 4 Asymptotic Properties of the QMLEs

In this section we study the consistency and asymptotic normality of the proposed QML estimators for the dynamic panel data models with spatial errors. We first state and discuss a set of generic assumptions applicable to all three scenarios discussed in Section 3. Then we proceed with each specific scenario where, under some additional assumptions, the key asymptotic results are presented. To facilitate the presentation, some general notation (old and new) is given.

**General notation:** (i) recall  $\psi = (\theta', \sigma_v^2, \delta')'$ , where  $\theta$  and  $\sigma_v^2$  are the linear and scale parameters and can be concentrated out from the likelihood function, and  $\delta$  is the vector of nonlinear parameters left in the concentrated likelihood function. Let  $\psi_0 = (\theta_0', \sigma_{v0}^2, \delta_0')'$  be the true parameter vector. Let  $\Psi$  be the parameter space of  $\psi$ , and  $\Delta$  the space of  $\delta$ . (ii) A parametric function, or vector, or matrix, evaluated at  $\psi_0$ , is denoted by adding a subscript  $_0$ , e.g.,  $B_0 = B|_{\lambda=\lambda_0}$ , and similarly for  $\Omega_0$ ,  $\Omega_0^*$ ,  $\Omega_0^\dagger$ , etc. (iii) The common expectation and variance operators ‘E’ and ‘Var’ correspond to  $\psi_0$ .

### 4.1 Generic assumptions

To provide a rigorous analysis of the QMLEs, we need to assume different sets of conditions based on different model specifications. Nevertheless, for both the random and fixed effects specifications we first make the following generic assumptions.

**Assumption G1:** (i) The available observations are:  $(y_{it}, x_{it}, z_i), i = 1, \dots, n, t = 0, 1, \dots, T$ , with  $T \geq 2$  fixed and  $n \rightarrow \infty$ ; (ii) The disturbance vector  $u_t = (u_{1t}, \dots, u_{nt})'$  exhibits both individual effects and spatially autocorrelated structure defined in (2.2) and (2.3) and  $v_{it}$  are iid for all  $i$  and  $t$  with  $E(v_{it}) = 0$ ,  $\text{Var}(v_{it}) = \sigma_v^2$ , and  $E|v_{it}|^{4+\epsilon_0} < \infty$  for some  $\epsilon_0 > 0$ ; (iii)  $\{x_{it}, t = \dots, -1, 0, 1, \dots\}$  and  $\{z_i\}$  are strictly exogenous and independent across  $i$ ; (iv)  $|\rho| < 1$  in (2.1); and (v) The true parameter  $\delta_0$  lies in the interior of  $\Delta$ , a convex compact set.

Assumption G1(i) corresponds to traditional panel data models with large  $n$  and small  $T$ . One can consider extending the QMLE procedure to panels with large  $n$  and large  $T$ ; see, for example, Phillips and Sul (2003). Assumption G1(ii) is standard in the literature. Assumption G1(iii) is not as strong as it appears in the spatial econometrics literature, since in most spatial analysis regressors are treated as being nonstochastic (e.g., Anselin, 1988; Kelejian and Prucha, 1998, 1999, 2010; Lee, 2004; Lin and Lee, 2010; Robinson, 2010; Su and Jin, 2010; Su, 2012). One can relax the strict exogeneity condition in Assumption G1(iii) like Hsiao et al. (2002) but this will complicate our analysis in case of spatially correlated errors. Assumption G1(iv) can be relaxed for the case of random effects with exogenous initial observations without any change of the derivation. It can also be relaxed for the fixed effects model with some modification of the derivation as in Hsiao et al. (2002). Assumption G1(v) is commonly assumed in the literature but deserves some further discussion.

For QML estimation, it is required that  $\lambda$  lies within a certain space to guarantee the non-singularity of  $I_n - \lambda W$ . If the eigenvalues of  $W$  are all real, then such a space is  $(w_{\min}^{-1}, w_{\max}^{-1})$  where  $w_{\min}$  and  $w_{\max}$  are, respectively, the smallest and the largest eigenvalues of  $W$ ; if, further,  $W$  is row normalized, then  $w_{\max} = 1$  and  $w_{\min}^{-1} < -1$ , and the parameter space of  $\lambda$  becomes  $(w_{\min}^{-1}, 1)$  (Anselin, 1988). In general, the eigenvalues of  $W$  may not be all real as  $W$  can be asymmetric. LeSage and Pace (2009, p. 88-89) argue that only the purely real eigenvalues can affect the singularity of  $I_n - \lambda W$ . Consequently, for  $W$

with complex eigenvalues, the interval of  $\lambda$  that guarantees non-singular  $I_n - \lambda W$  is  $(w_s^{-1}, 1)$  where  $w_s$  is the most negative real eigenvalue of  $W$ . Kelejian and Prucha (2010) suggest the parameter space be  $(-\tau_n^{-1}, \tau_n^{-1})$  where  $\tau_n$  is the spectral radius of  $W$ , which is normalized to  $(-1, 1)$  by a single factor  $\tau_n^{-1}$ .

For the spatial weight matrix, we make the following assumptions.

**Assumption G2:** (i) The elements  $w_{ij}$  of  $W$  are at most of order  $h_n^{-1}$ , denoted by  $O(h_n^{-1})$ , uniformly in all  $i$  and  $j$ . As a normalization,  $w_{ii} = 0$  for all  $i$ ; (ii) The ratio  $h_n/n \rightarrow 0$  as  $n$  goes to infinity; (iii) The matrix  $B_0$  is nonsingular; (iv) The sequences of matrices  $\{W\}$  and  $\{B_0^{-1}\}$  are uniformly bounded in both row and column sums; (v)  $\{B^{-1}\}$  are uniformly bounded in either row or column sums, uniformly in  $\lambda$  in a compact parameter space  $\Lambda$ , and  $\underline{c}_\lambda \leq \inf_{\lambda \in \Lambda} \lambda_{\max}(B' B) \leq \sup_{\lambda \in \Lambda} \lambda_{\max}(B' B) \leq \bar{c}_\lambda < \infty$ .

Assumptions G2(i)-(iv) parallel Assumptions 2-4 of Lee (2004). Like Lee (2004), Assumptions G2(i)-(iv) provide the essential features of the weight matrix for the model. Assumption G2(ii) is always satisfied if  $\{h_n\}$  is a bounded sequence. We allow  $\{h_n\}$  to be divergent but at a rate smaller than  $n$  as in Lee (2004). Assumption G2(iii) guarantees that the disturbance term is well defined. Kelejian and Prucha (1998, 1999, 2001) and Lee (2004) also assume Assumption G2(iv) which limits the spatial correlation to some degree but facilitates the study of the asymptotic properties of the spatial parameter estimators. By Horn and Johnson (1985, p. 301), that  $\limsup_n \|\lambda_0 W\| < 1$  guarantees that  $B_0^{-1}$  is uniformly bounded in both row and column sums. By Lee (2002, Lemma A.3), Assumption G2(iv) implies  $\{B^{-1}\}$  are uniformly bounded in both row and column sums uniformly in a neighborhood of  $\lambda_0$ . Assumption G2(v) is stronger than Assumption G2(iv) and is required in establishing the consistency results.

## 4.2 Random effects model

We now present detailed asymptotic results for the SDPD model with random effects. Beside the generic assumptions given earlier, some additional assumptions specific for this model are necessary.

**Assumption R:** (i)  $\mu_i$ 's are iid with  $E(\mu_i) = 0$ ,  $Var(\mu_i) = \sigma_\mu^2$ , and  $E|\mu_i|^{4+\epsilon_0} < \infty$  for some  $\epsilon_0 > 0$ ; (ii)  $\mu_i$  and  $v_{jt}$  are mutually independent, and they are independent of  $x_{ks}$  and  $z_k$  for all  $i, j, k, t, s$ ; (iii) All elements in  $(x_{it}, z_i)$  have  $4 + \epsilon_0$  moments for some  $\epsilon_0 > 0$ .

Assumption R(i) and the first part of Assumption R(ii) are standard in the random effects panel data literature. The second part of Assumption R(ii) is for convenience. Alternatively we can treat the regressors as being nonstochastic.

**Case I:  $y_0$  is exogenous.** To derive the consistency of the QML estimators, we need to ensure that  $\delta = (\lambda, \phi_\mu)'$  is identifiable. Then, the identifiability of other parameters follows. Following White (1994) and Lee (2004), define  $\mathcal{L}_c^{r*}(\delta) = \max_{\theta, \sigma_v^2} E[\mathcal{L}^r(\theta, \sigma_v^2, \delta)]$ , where we suppress the dependence of  $\mathcal{L}_c^{r*}(\delta)$  on  $n$ . The optimal solution to  $\max_{\theta, \sigma_v^2} E[\mathcal{L}^r(\theta, \sigma_v^2, \delta)]$  is given by

$$\tilde{\theta}(\delta) = [E(\tilde{X}' \Omega^{-1} \tilde{X})]^{-1} E(\tilde{X}' \Omega^{-1} Y) \text{ and} \quad (4.1)$$

$$\tilde{\sigma}_v^2(\delta) = \frac{1}{nT} E[u(\tilde{\theta}(\delta))' \Omega^{-1} u(\tilde{\theta}(\delta))]. \quad (4.2)$$

Consequently, we have

$$\mathcal{L}_c^{r*}(\delta) = -\frac{nT}{2} [\log(2\pi) + 1] - \frac{nT}{2} \log[\tilde{\sigma}_v^2(\delta)] - \frac{1}{2} \log |\Omega|. \quad (4.3)$$

Noting that  $\tilde{\theta}(\delta_0) = \theta_0 + [E(\tilde{X}'\Omega_0^{-1}\tilde{X})]^{-1}E(\tilde{X}'\Omega_0^{-1}u) = \theta_0$  by Lemma B.6, we can readily show that  $\tilde{\sigma}_v^2(\delta_0) = \sigma_{v0}^2$ . We impose the following identification condition.

**Assumption R:** (iv)  $\lim_{n \rightarrow \infty} \frac{1}{nT} \{ \log |\sigma_{v0}^2 \Omega_0| - \log |\tilde{\sigma}_v^2(\delta) \Omega(\delta)| \} \neq 0$  for any  $\delta \neq \delta_0$ , and  $\frac{1}{nT} \tilde{X}'\tilde{X}$  is positive definite almost surely for sufficiently large  $n$ .

The first part of Assumption R(iv) parallels Assumption 9 in Lee (2004). It is a global identification condition related to the uniqueness of the variance-covariance matrix of  $u$ . With this and the uniform convergence of  $\frac{1}{nT}[\mathcal{L}_c^r(\delta) - \mathcal{L}_c^{r*}(\delta)]$  to zero on  $\Delta$  proved in the Appendix C, the consistency of  $\hat{\delta}$  follows. The consistency of  $\hat{\theta}$  and  $\hat{\sigma}_v^2$  follows from that of  $\hat{\delta}$  and the second part of Assumption R(iv).

**Theorem 4.1** Under Assumptions G1, G2, and R(i)-(iv), if the initial observations  $y_{i0}$  are exogenously given, then  $\hat{\psi} \xrightarrow{p} \psi_0$ .

To derive the asymptotic distribution of  $\hat{\psi}$ , we need to make a Taylor expansion of  $\frac{\partial}{\partial \psi} \mathcal{L}^r(\hat{\psi}) = 0$  at  $\psi_0$ , and then to check that the score function and Hessian matrix have proper asymptotic behavior. First, the score function  $S^r(\psi) = \frac{\partial}{\partial \psi} \mathcal{L}^r(\psi)$  has the elements

$$\begin{aligned} \frac{\partial \mathcal{L}^r(\psi)}{\partial \theta} &= \frac{1}{\sigma_v^2} \tilde{X}' \Omega^{-1} u(\theta), \\ \frac{\partial \mathcal{L}^r(\psi)}{\partial \sigma_v^2} &= \frac{1}{2\sigma_v^4} u(\theta)' \Omega^{-1} u(\theta) - \frac{nT}{2\sigma_v^2}, \\ \frac{\partial \mathcal{L}^r(\psi)}{\partial \omega} &= \frac{1}{2\sigma_v^2} u(\theta)' P_\omega u(\theta) - \frac{1}{2} \text{tr}(P_\omega \Omega), \quad \omega = \lambda, \phi_\mu, \end{aligned}$$

where  $P_\omega = \Omega^{-1} \Omega_\omega \Omega^{-1}$  and  $\Omega_\omega = \frac{\partial}{\partial \omega} \Omega(\delta)$  for  $\omega = \lambda, \phi_\mu$ . One can easily verify that  $\Omega_\lambda = I_T \otimes A$  and  $\Omega_{\phi_\mu} = J_T \otimes I_n$  where  $A = \frac{\partial}{\partial \lambda} (B'B)^{-1} = (B'B)^{-1} (W'B + B'W) (B'B)^{-1}$ . At  $\psi = \psi_0$ , the last three components of the score function are linear and quadratic functions of  $u \equiv u(\theta_0)$  and one can readily verify that their expectations are zero. The first score component contains  $\frac{1}{\sigma_v^2} Y_{-1}' \Omega^{-1} u(\theta)$ , and some additional algebra is needed to prove  $E[Y_{-1}' \Omega_0^{-1} u(\theta_0)] = 0$ , which is given in Lemma B.6.

Asymptotic normality of the score, proved in Lemma B.8, is essential for the asymptotic normality of the QMLEs. Note that the elements in  $u$  are not independent and that  $\tilde{X}$  contains the lagged dependent variable  $Y_{-1}$ , thus the standard results, such as the central limit theorem (CLT) for linear and quadratic forms in Kelejian and Prucha (2001) cannot be directly applied. For the last three components, we need to plug  $u = (\iota_T \otimes I_n)\mu + (I_T \otimes B_0^{-1})v$  into  $S^r(\psi_0)$  and apply the CLT to linear and quadratic functions of  $\mu$  and  $v$  separately. For the first component, a special care has to be given to  $Y_{-1}$  (see Lemma B.8).

Let  $H_{r,n}(\psi) \equiv \frac{\partial^2}{\partial \psi \partial \psi'} \mathcal{L}^r(\psi)$  be the Hessian matrix, and  $\Gamma_{r,n}(\psi) = E[\frac{\partial}{\partial \psi} \mathcal{L}^r(\psi) \frac{\partial}{\partial \psi'} \mathcal{L}^r(\psi)]$  be the VC matrix of the score vector, both are given in Appendix A. Lemma B.7 shows that  $\frac{1}{nT} [H_{r,n}(\psi_0) - E H_{r,n}(\psi_0)] = o_p(1)$ . The asymptotic normality of the QMLE thus follows from the mean value theorem:  $0 = \frac{1}{\sqrt{nT}} S^r(\hat{\psi}) = \frac{1}{\sqrt{nT}} S^r(\psi_0) + \frac{1}{nT} H_{r,n}(\bar{\psi}) \cdot \sqrt{nT}(\hat{\psi} - \psi_0)$ , provided that  $\frac{1}{nT} [H_{r,n}(\bar{\psi}) - H_{r,n}(\psi_0)] = o_p(1)$  where  $\bar{\psi}$  lies between  $\hat{\psi}$  and  $\psi_0$  (see Appendix C for details). We have the following theorem.

**Theorem 4.2** Under Assumptions G1, G2, and R(i)-(iv), if the initial observations  $y_0$  are exogenously given, then

$$\sqrt{nT}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, H_r^{-1} \Gamma_r H_r^{-1}),$$

where  $H_r = \lim_{n \rightarrow \infty} \frac{1}{nT} E[H_{r,n}(\psi_0)]$  and  $\Gamma_r = \lim_{n \rightarrow \infty} \frac{1}{nT} \Gamma_{r,n}(\psi_0)$ , both assumed to exist, and  $(-H_r)$  is assumed to be positive definite. When errors are normally distributed,  $\sqrt{nT}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, H_r^{-1})$ .



As in Lee (2004), the asymptotic results in Theorem 4.2 is valid regardless of whether the sequence  $\{h_n\}$  is bounded or divergent. The matrices  $\Gamma_r$  and  $H_r$  can be simplified if  $h_n \rightarrow \infty$  as  $n \rightarrow \infty$ .<sup>10</sup>

**Case II:  $y_0$  is endogenous.** In this case, define  $\mathcal{L}_c^{rr*}(\delta) = \max_{\theta, \sigma_v^2} E[\mathcal{L}^{rr}(\theta, \sigma_v^2, \delta)]$ , where we suppress the dependence of  $\mathcal{L}_c^{rr*}(\delta)$  on  $n$ . The optimal solution to  $\max_{\theta, \sigma_v^2} E[\mathcal{L}^{rr}(\theta, \sigma_v^2, \delta)]$  is now given by

$$\tilde{\theta}(\delta) = [E(X^{*'}\Omega^{*-1}(\delta)X^*)]^{-1}E[X^{*'}\Omega^{*-1}(\delta)Y^*(\rho)], \text{ and} \quad (4.4)$$

$$\tilde{\sigma}_v^2(\delta) = \frac{1}{n(T+1)}E[u^*(\tilde{\theta}(\delta), \rho)' \Omega^{*-1}(\delta)u^*(\tilde{\theta}(\delta), \rho)]. \quad (4.5)$$

Consequently, we have

$$\mathcal{L}_c^{rr*}(\delta) = -\frac{n(T+1)}{2} [\log(2\pi) + 1] - \frac{n(T+1)}{2} \log \tilde{\sigma}_v^2(\delta) - \frac{1}{2} \log |\Omega^*|. \quad (4.6)$$

We make the following identification assumption.

**Assumption R:**  $(iv^*) \lim_{n \rightarrow \infty} \frac{1}{n(T+1)} \{ \log |\sigma_{v0}^2 \Omega_0^*| - \log |\tilde{\sigma}_v^2(\delta) \Omega^*(\delta)| \} \neq 0$  for any  $\delta \neq \delta_0$ . Both  $\frac{1}{n} \tilde{\mathbf{x}}' \tilde{\mathbf{x}}$  and  $\frac{1}{nT} (X, Z)'(X, Z)$  are positive definite almost surely for sufficiently large  $n$ .

The following theorem establishes the consistency of QMLE for the random effects model with endogenous initial observations. Similarly, the key result is to show that  $\frac{1}{n(T+1)}[\mathcal{L}_c^{rr}(\delta) - \mathcal{L}_c^{rr*}(\delta)]$  converges to zero uniformly in  $\delta \in \Delta$ , which is given in Appendix C.

**Theorem 4.3** Under Assumptions G1, G2, R0, R(i)-(iii) and R(iv\*), if the initial observations  $y_0$  are endogenously given, then  $\hat{\psi} \xrightarrow{P} \psi_0$ .

Again, to derive the asymptotic distribution of  $\hat{\psi}$ , one starts with a Taylor expansion of the score function,  $S^{rr}(\psi) = \frac{\partial}{\partial \psi} \mathcal{L}^{rr}(\psi)$ , of which the elements are given below:

$$\begin{aligned} \frac{\partial \mathcal{L}^{rr}(\psi)}{\partial \theta} &= \frac{1}{\sigma_v^2} X^{*'} \Omega^{*-1} u^*(\theta, \rho), \\ \frac{\partial \mathcal{L}^{rr}(\psi)}{\partial \sigma_v^2} &= \frac{1}{2\sigma_v^4} u^*(\theta, \rho)' \Omega^{*-1} u^*(\theta, \rho) - \frac{n(T+1)}{2\sigma_v^2}, \\ \frac{\partial \mathcal{L}^{rr}(\psi)}{\partial \rho} &= -\frac{1}{\sigma_v^2} u^*(\theta, \rho)' \Omega^{*-1} u^*(\theta, \rho) + \frac{1}{2\sigma_v^2} u^*(\theta, \rho)' P_\rho^* u^*(\theta, \rho) - \frac{1}{2} \text{tr}(P_\rho^* \Omega^*), \\ \frac{\partial \mathcal{L}^{rr}(\psi)}{\partial \omega} &= \frac{1}{2\sigma_v^2} u^*(\theta, \rho)' P_\omega^* u^*(\theta, \rho) - \frac{1}{2} \text{tr}(P_\omega^* \Omega^*), \text{ for } \omega = \lambda, \phi_\mu, \text{ and } \phi_\zeta, \end{aligned}$$

where  $u_\rho^*(\theta, \rho) = \frac{\partial}{\partial \rho} u^*(\theta, \rho)$ ,  $P_\omega^* = \Omega^{*-1} \Omega_\omega^* \Omega^{*-1}$ , and  $\Omega_\omega^* = \frac{\partial}{\partial \omega} \Omega^*(\delta)$  for  $\omega = \rho, \lambda, \phi_\mu$ , and  $\phi_\zeta$ , given as

$$\begin{aligned} u_\rho^*(\theta, \rho) &= - \begin{pmatrix} \dot{a}_m Z \gamma \\ Y_{-1} \end{pmatrix}, \quad \Omega_\rho^* = \begin{pmatrix} 2\phi_\mu \dot{a}_m \dot{a}_m' I_n + \dot{b}_m (B' B)^{-1} & \phi_\mu \dot{a}_m (\iota' \otimes I_n) \\ \phi_\mu \dot{a}_m (\iota \otimes I_n) & 0_{nT \times nT} \end{pmatrix}, \\ \Omega_\lambda^* &= \begin{pmatrix} b_m & 0_T' \\ 0_T & I_T \end{pmatrix} \otimes A, \quad \Omega_{\phi_\mu}^* = \begin{pmatrix} a_m^2 & a_m \iota_T' \\ a_m \iota_T & J_T \end{pmatrix} \otimes I_n, \text{ and } \Omega_{\phi_\zeta}^* = \begin{pmatrix} 1 & 0_T' \\ 0_T & 0_{T \times T} \end{pmatrix} \otimes I_n, \end{aligned}$$

where  $\dot{a}_m = \frac{d}{d\rho} a_m(\rho)$  and  $\dot{b}_m = \frac{d}{d\rho} b_m(\rho)$ , and their expressions can easily be obtained. One can readily verify that  $E[\frac{\partial}{\partial \psi} \mathcal{L}^{rr}(\psi_0)] = 0$ . The asymptotic normality of the score is given in Lemma B.13. The

<sup>10</sup>It can be shown, by some algebra similar to these for proving Lemma B.6 but using (B.2) instead of (B.3), that when  $T$  is also large the results of Theorems 4.1 and 4.2 remain valid under an endogenous  $y_0$ , although issues such as the exact rate of convergence and the magnitude of bias remain. Nevertheless, it shows that when  $T$  is also large one can indeed ignore the endogeneity of  $y_0$ , as it was done in Yu et al. (2008) for a fixed effects spatial lag SDPD model with both large  $n$  and large  $T$ . However, a detailed study along this line is clearly beyond the scope of the paper.

asymptotic normality of the QMLE thus follows if the Hessian matrix,  $H_{rr,n}(\psi) \equiv \frac{\partial^2}{\partial \psi \partial \psi'} \mathcal{L}^{rr}(\psi)$ , given in Appendix A, possesses the desired stochastic convergence property as those for the case of exogenous  $y_0$ .

Let  $\Gamma_{rr,n}(\psi) = E[\frac{\partial}{\partial \psi} \mathcal{L}^{rr}(\psi) \frac{\partial}{\partial \psi'} \mathcal{L}^{rr}(\psi)]$  be the variance-covariance matrix of the score vector with its detail given in Appendix A. We now state the asymptotic normality result.

**Theorem 4.4** *Under Assumptions G1, G2, R0, R(i)-(iii) and R(iv\*), if the initial observations are endogenously given, then*

$$\sqrt{nT}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, H_{rr}^{-1} \Gamma_{rr} H_{rr}^{-1}),$$

where  $H_{rr} = \lim_{n \rightarrow \infty} \frac{1}{n(T+1)} E[H_{rr,n}(\psi_0)]$  and  $\Gamma_{rr} = \lim_{n \rightarrow \infty} \frac{1}{n(T+1)} \Gamma_{rr,n}(\psi_0)$ , both assumed to exist, and  $(-H_{rr})$  is assumed to be positive definite. When errors are normal,  $\sqrt{nT}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, H_{rr}^{-1})$ .

### 4.3 Fixed effects model

For the fixed effects model, we need to supplement the generic assumptions, Assumptions G1 and G2, made above with the following assumption on the regressors.

**Assumption F:** (i) The processes  $\{x_{it}, t = \dots, -1, 0, 1, \dots\}$  are trend-stationary or first-differencing stationary for all  $i = 1, \dots, n$ ; (ii) All elements in  $\Delta x_{it}$  have  $4 + \epsilon_0$  moments for some  $\epsilon_0 > 0$ ; (iii)  $\frac{1}{nT} \Delta X^\dagger \Delta X^\dagger$  is positive definite almost surely for sufficiently large  $n$ .

Define  $\mathcal{L}_c^{f*}(\delta) = \max_{\theta, \sigma_v^2} E[\mathcal{L}^f(\theta, \sigma_v^2, \delta)]$ , where we suppress the dependence of  $\mathcal{L}_c^{f*}(\delta)$  on  $n$ . The optimal solution to  $\max_{\theta, \sigma_v^2} E[\mathcal{L}^f(\theta, \sigma_v^2, \delta)]$  is now given by

$$\tilde{\theta}(\delta) = \{E[(\Delta X^\dagger)' \Omega^{\dagger-1} \Delta X^\dagger]\}^{-1} E[(\Delta X^\dagger)' \Omega^{\dagger-1} \Delta Y^\dagger(\rho)] \text{ and} \quad (4.7)$$

$$\tilde{\sigma}_v^2(\delta) = \frac{1}{nT} E[\Delta u(\tilde{\theta}(\delta), \rho)' \Omega^{\dagger-1} \Delta u(\tilde{\theta}(\delta), \rho)]. \quad (4.8)$$

Consequently, we have

$$\mathcal{L}_c^{f*}(\delta) = -\frac{nT}{2} [\log(2\pi) + 1] - \frac{nT}{2} \log[\tilde{\sigma}_v^2(\delta)] - \frac{1}{2} \log |\Omega^\dagger|. \quad (4.9)$$

The following identification condition is needed for our consistency result.

**Assumption F:** (iv)  $\lim_{n \rightarrow \infty} \frac{1}{nT} \{ \log |\sigma_{v0}^2 \Omega_0^\dagger| - \log |\tilde{\sigma}_v^2(\delta) \Omega^\dagger(\delta)| \} \neq 0$  for any  $\delta \neq \delta_0$ .

With this identification condition, the consistency of  $\hat{\delta}$  follows if  $\frac{1}{nT} [\mathcal{L}_c^f(\delta) - \mathcal{L}_c^{f*}(\delta)]$  converges to zero uniformly on  $\Delta$ . The consistency of  $\hat{\theta}$  and  $\hat{\sigma}_v^2$  then follows from the consistency of  $\hat{\delta}$  and the identification condition given in Assumption F(iii). We have the following theorem.

**Theorem 4.5** *Under Assumptions G1, G2, F0, and F, we have for either exogenous or endogenous  $y_0$ ,  $\hat{\psi} \xrightarrow{P} \psi_0$ .*

To derive the asymptotic distribution of  $\hat{\psi}$ , one needs the score function  $S^f(\psi) = \frac{\partial}{\partial \psi} \mathcal{L}^f(\psi)$ :

$$\begin{aligned} \frac{\partial \mathcal{L}^f(\psi)}{\partial \theta} &= \frac{1}{\sigma_v^2} \Delta X^\dagger \Omega^{\dagger-1} \Delta u(\theta, \rho), \\ \frac{\partial \mathcal{L}^f(\psi)}{\partial \sigma_v^2} &= \frac{1}{2\sigma_v^4} \Delta u(\theta, \rho)' \Omega^{\dagger-1} \Delta u(\theta, \rho) - \frac{nT}{2\sigma_v^2}, \\ \frac{\partial \mathcal{L}^f(\psi)}{\partial \rho} &= -\frac{1}{\sigma_v^2} \Delta u_\rho(\theta, \rho)' \Omega^{\dagger-1} \Delta u(\theta, \rho) + \frac{1}{2\sigma_v^2} \Delta u(\theta, \rho)' P_\rho^\dagger \Delta u(\theta, \rho) - \frac{1}{2} \text{tr}(\Omega^{\dagger-1} \Omega_\rho^\dagger), \\ \frac{\partial \mathcal{L}^f(\psi)}{\partial \omega} &= \frac{1}{2\sigma_v^2} \Delta u(\theta, \rho)' P_\omega^\dagger \Delta u(\theta, \rho) - \frac{1}{2} \text{tr}(\Omega^{\dagger-1} \Omega_\omega^\dagger) \text{ for } \omega = \lambda, \phi_\zeta, \end{aligned}$$

where  $\Delta u_\rho(\theta, \rho) = \frac{\partial}{\partial \rho} \Delta u(\theta, \rho) = -(0'_{n \times 1}, \Delta y'_1, \dots, \Delta y'_{T-1})'$ , and  $\Omega_\omega^\dagger = \frac{\partial}{\partial \omega} \Omega^\dagger(\delta)$  and  $P_\omega^\dagger = \Omega^{\dagger-1} \Omega_\omega^\dagger \Omega^{\dagger-1}$  for  $\omega = \rho, \lambda$ , and  $\phi_\zeta$ . From (3.24), it is easy to see that  $\Omega_\rho^\dagger = h_{\dot{c}_m} \otimes (B'B)^{-1}$ ,  $\Omega_\lambda^\dagger = h_{c_m} \otimes A$ , and  $\Omega_{\phi_\zeta}^\dagger = \ell_1 \otimes I_n$ , where  $\dot{c}_m = \frac{\partial}{\partial \rho} c_m(\rho)$ . Again, one can readily verify that  $E[\frac{\partial}{\partial \psi} \mathcal{L}^f(\psi_0)] = 0$ . The asymptotic normality of the score is given in Lemma B.15. The asymptotic normality of  $\hat{\psi}$  thus follows if the Hessian matrix,  $H_{f,n}(\psi) \equiv \frac{\partial^2}{\partial \psi \partial \psi'} \mathcal{L}^f(\psi)$ , given in Appendix A, possesses the desired stochastic convergence properties as those for random effects model. Let  $\Gamma_{f,n}(\psi) = E[\frac{\partial}{\partial \psi} \mathcal{L}^f(\psi) \frac{\partial}{\partial \psi'} \mathcal{L}^f(\psi)]$  be the VC matrix of the score vector, given in Appendix A. We now state the asymptotic normality result.

**Theorem 4.6** *Under Assumptions G1, G2, F0 and F, we have for either exogenous or endogenous  $y_0$ ,*

$$\sqrt{nT}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, H_f^{-1} \Gamma_f H_f^{-1}),$$

where  $H_f = \lim_{n \rightarrow \infty} \frac{1}{nT} E[H_{f,n}(\psi_0)]$  and  $\Gamma_f = \lim_{n \rightarrow \infty} \frac{1}{nT} \Gamma_{f,n}(\psi_0)$ , both assumed to exist, and  $(-H_f)$  is assumed to be positive definite. When errors are normally distributed,  $\sqrt{nT}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, H_f^{-1})$ .

## 5 Bootstrap Estimate of the Variance-Covariance Matrix

From Theorems 4.2, 4.4 and 4.6, we see that the asymptotic variance-covariance (VC) matrices of the QMLEs of the three models considered are, respectively,  $H_r^{-1} \Gamma_r H_r^{-1}$ ,  $H_{rr}^{-1} \Gamma_{rr} H_{rr}^{-1}$ , and  $H_f^{-1} \Gamma_f H_f^{-1}$ . Practical applications of the asymptotic normality theory depend upon the availability of a consistent estimator of the asymptotic VC matrix. Obviously, the Hessian matrices evaluated at the QMLEs provide consistent estimators for  $H_r, H_{rr}$ , and  $H_f$ , i.e.,  $\hat{H}_r \equiv \frac{1}{nT} H_{r,n}(\hat{\psi})$ ,  $\hat{H}_{rr} \equiv \frac{1}{n(T+1)} H_{rr,n}(\hat{\psi})$ , and  $\hat{H}_f \equiv \frac{1}{nT} H_{f,n}(\hat{\psi})$ . The formal proofs of the consistency of these estimators can be found in the proofs of Theorems 4.2, 4.4, and 4.6, respectively. However, consistent estimators for  $\Gamma_r, \Gamma_{rr}$ , and  $\Gamma_f$ , the asymptotic VC matrices of the scores (normalized), are not readily available due to the presence of error components in the original model and in the model for the initial observations.<sup>11</sup>

As indicated in the introduction, the traditional methods based on sample-analogues, outer product of gradients (OPG), or closed form expressions, do not provide an easy solution. First, from the expressions of the score functions,  $S^r(\psi), S^{rr}(\psi)$  and  $S^f(\psi)$ , given in Section 4, we see that it is very difficult, if possible at all, to find sample analogues of  $E[S^r(\psi_0)S^r(\psi_0)']$ ,  $E[S^{rr}(\psi_0)S^{rr}(\psi_0)']$  and  $E[S^f(\psi_0)S^f(\psi_0)']$ , bearing in mind that  $S^r(\hat{\psi}), S^{rr}(\hat{\psi})$  and  $S^f(\hat{\psi})$  are all zero by the definition of the QMLEs. Second, OPG method typically requires that the score function be written as a single summation of  $n$  uncorrelated terms. This cannot be done in our framework as our score function has the form of a second order  $V$  statistic instead. Third, although the closed form expressions for the VC matrices can be derived (see Appendix D), these expressions typically contain the third and fourth moments of the error components in models (3.1), (3.10), (3.15) and (3.19). Some elements of the VC matrices cannot be consistently estimated due to the complicated interaction of the error terms with the lagged dependent variable and the fact that only a short panel data is available. Thus, an alternative method is desired.

In this section, we introduce a residual-based bootstrap method for estimating the VC matrices of the scores, with the bootstrap draws made on the joint empirical distribution function (EDF) of  $n$

<sup>11</sup>This is not a problem for the exact likelihood inference (Elhorst, 2005, Yang et al., 2006) as in this case the VC matrix of the score function equals the negative expected Hessian. Hence, the asymptotic VC matrices of the MLEs in the three models considered reduce to  $-H_r^{-1}$ ,  $-H_{rr}^{-1}$  and  $-H_f^{-1}$ , respectively, of which sample analogues exist.

transformed vectors of residuals. While the general principle for our bootstrap method is the same for all the three models considered above, different structures of the residuals and the score functions render them a separate consideration.

## 5.1 Random effects model with exogenous initial values

Write the model as:  $y_t = \rho_0 y_{t-1} + x_t \beta_0 + z \gamma_0 + u_t$ ,  $u_t = \mu + B_0^{-1} v_t$ ,  $t = 1, 2, \dots, T$ , now viewed as a real-world data generating process (DGP). We have,  $\text{Var}(u_t) = \sigma_{v0}^2 (\phi_{\mu 0} I_n + (B_0' B_0)^{-1}) \equiv \sigma_{v0}^2 \Sigma(\lambda_0, \phi_{\mu 0})$ . Define the *transformed residuals* ( $t$ -residuals):

$$r_t = \Sigma^{-\frac{1}{2}}(\lambda_0, \phi_{\mu 0}) u_t, \quad t = 1, \dots, T,$$

where  $\Sigma^{\frac{1}{2}}(\lambda, \phi_{\mu})$  is a square-root matrix of  $\Sigma(\lambda, \phi_{\mu})$ . Then,  $E(r_t) = 0$  and  $\text{Var}(r_t) = \sigma_{v0}^2 I_n$ . Thus, the elements of  $r_t$  are uncorrelated, which are iid if  $\mu$  and  $v_t$  are normal satisfying the conditions given in Assumptions G1 and R. As our asymptotics depend only on  $n$ , these uncorrelated errors lay out the theoretical foundation for a residual-based bootstrap method. Let  $\hat{r}_t$  be the QML estimate of  $r_t$ , and  $\hat{\mathcal{F}}_{n,t}$  be the empirical distribution function (EDF) of the centered  $\hat{r}_t$ , for  $t = 1, 2, \dots, T$ . Let  $S^r(Y_{-1}, u, \psi_0)$  be the score function given below Theorem 4.1, now written in terms of the lagged response  $Y_{-1}$ , the disturbance vector  $u$  and the true parameter vector  $\psi_0$ . The bootstrap procedure for estimating  $\Gamma_{n,r}(\psi_0)$  is as follows.

1. Compute the QMLE  $\hat{\psi}$ , the QML residuals  $\hat{u}_t = y_t - \hat{\rho} y_{t-1} - x_t \hat{\beta} - z \hat{\gamma}$ , and the transformed QML residuals  $\hat{r}_t = \Sigma^{-\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_{\mu}) \hat{u}_t$ , for  $t = 1, 2, \dots, T$ . For each  $t$ , center  $\hat{r}_t$  by its mean.
2. Make  $n$  random draws from the rows of  $(\hat{r}_1, \dots, \hat{r}_T)$  to give  $T$  *matched* bootstrap samples,  $\{\hat{r}_1^b, \dots, \hat{r}_T^b\}$ , of the transformed residuals.
3. Conditional on  $y_0, \mathbf{x}, z$ , and the QMLE  $\hat{\psi}$ , generate the bootstrap data according to

$$\begin{aligned} \hat{y}_1^b &= \hat{\rho} y_0 + x_1 \hat{\beta} + z \hat{\gamma} + \Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_{\mu}) \hat{r}_1^b, \\ \hat{y}_t^b &= \hat{\rho} \hat{y}_{t-1}^b + x_t \hat{\beta} + z \hat{\gamma} + \Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_{\mu}) \hat{r}_t^b, \quad t = 2, 3, \dots, T. \end{aligned}$$

The bootstrapped values of  $u$  and  $Y_{-1}$  are given by  $\hat{u}^b = \text{vec}[\Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_{\mu})(\hat{r}_1^b, \dots, \hat{r}_T^b)]$  and  $\hat{Y}_{-1}^b = \text{vec}(y_0, \hat{y}_1^b, \dots, \hat{y}_{T-1}^b)$ , respectively.

4. Compute  $S^r(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi})$ , the score function in the bootstrap world.
5. Repeat steps 2-4  $B$  times, and the bootstrap estimate of  $\Gamma_{n,r}(\psi_0)$  is given by

$$\hat{\Gamma}_{n,r}^b = \frac{1}{B} \sum_{b=1}^B S^r(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi}) S^r(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi})' - \frac{1}{B} \sum_{b=1}^B S^r(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi}) \cdot \frac{1}{B} \sum_{b=1}^B S^r(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi})'. \quad (5.1)$$

An intuitive justification for the validity of the above bootstrap procedure goes as follows. First, note that the score function can be written as  $S^r(Y_{-1}, u, \psi_0)$ , viewed as a function of random components and parameters. Note that  $u_t = \mu + B_0^{-1} v_t$ ,  $t = 1, \dots, T$ . If  $\psi_0$  and the distributions of  $\mu_i$  and  $v_{it}$  were all known, then to compute the value of  $\Gamma_{n,r}(\psi_0)$ , one can simply use the Monte Carlo method: (i) generate

Monte Carlo samples  $\mu^m$  and  $v_t^m, t = 1, \dots, T$ , to give a Monte Carlo value  $u^m$ , (ii) compute the Monte Carlo value  $Y_{-1}^m$  based on  $u^m, \{x_1, \dots, x_T\}$ , and  $z$ , through the real-world DGP, (iii) compute a Monte Carlo value  $S^{r,m}(\psi_0) = S^r(Y_{-1}^m, u^m, \psi_0)$  for the score function, and (iv) repeat (i)-(iii)  $M$  times to give a Monte Carlo approximation to the value of  $\Gamma_{n,r}(\psi_0)$  as

$$\Gamma_{n,r}^m(\psi_0) \approx \frac{1}{M} \sum_{m=1}^B S^{r,m}(\psi_0) S^{r,m}(\psi_0)' - \frac{1}{M} \sum_{m=1}^M S^{r,m}(\psi_0) \cdot \frac{1}{M} \sum_{m=1}^M S^{r,m}(\psi_0)', \quad (5.2)$$

which can be made to an arbitrary level of accuracy by choosing an arbitrarily large  $M$ . Note that  $u_t = \Sigma^{\frac{1}{2}}(\lambda_0, \phi_{\mu 0}) r_t$ . The step (i) above is equivalent to draw random sample  $r_t^m$  from the joint distribution  $\mathcal{F}_t$  of  $r_t$ , and compute  $u_t^m = \Sigma^{\frac{1}{2}}(\lambda_0, \phi_{\mu 0}) r_t^m$ .<sup>12</sup>

However, in the real world,  $\psi_0$  is unknown. In this case, it is clear that a Monte Carlo estimate of  $\Gamma_{n,r}(\psi_0)$  can be obtained by plugging  $\hat{\psi}$  into (5.2),

$$\hat{\Gamma}_{n,r}^m = \frac{1}{M} \sum_{m=1}^B S^{r,m}(\hat{\psi}) S^{r,m}(\hat{\psi})' - \frac{1}{M} \sum_{m=1}^M S^{r,m}(\hat{\psi}) \cdot \frac{1}{M} \sum_{m=1}^M S^{r,m}(\hat{\psi})'. \quad (5.3)$$

In the real world the distributions of  $\mu_i$  and  $v_{it}$ , and hence  $\mathcal{F}_t$ , are also unknown. However, we note that the only difference between  $\hat{\Gamma}_{n,r}^b$  given in (5.1) and  $\hat{\Gamma}_{n,r}^m$  given in (5.3) is that  $\hat{r}_t^b$  for the former is from the EDF  $\hat{\mathcal{F}}_{n,t}$ , but  $r_t^m$  for the latter is drawn from the true joint distribution  $\mathcal{F}_t$ . The bootstrap DGP that mimics the real-world DGP must be  $\hat{y}_1^b = \hat{\rho} y_0 + x_1 \hat{\beta} + z \hat{\gamma} + \hat{u}_1^b$ , and  $\hat{y}_t^b = \hat{\rho} \hat{y}_{t-1}^b + x_t \hat{\beta} + z \hat{\gamma} + \hat{u}_t^b, t = 2, \dots, T$ . Thus, if  $\{\hat{\mathcal{F}}_{n,t}\}$  are able to produce  $\{\hat{r}_t^b\}$  that mimic  $\{r_t^m\}$  drawn from  $\{\mathcal{F}_t\}$  up to the fourth moments, which is typically the case as  $\hat{\psi}$  is consistent for  $\psi_0$  and the spatial weight matrix is typically sparse (see Appendix D for details), then  $\hat{\Gamma}_{n,r}^b$  and  $\hat{\Gamma}_{n,r}^m$  are asymptotically equivalent. The extra variability caused by replacing  $\mathcal{F}_t$  by  $\hat{\mathcal{F}}_{n,t}$  is of the same order as that from replacing  $\psi_0$  by  $\hat{\psi}$ . This justifies the validity of the proposed bootstrap procedure in a heuristic manner.

Formally, let  $\text{Var}^b(S^r(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi}))$  be the true bootstrap variance of  $S^r(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi})$  where the variance operator  $\text{Var}^b$  corresponds to the EDFs  $\{\hat{\mathcal{F}}_{n,t}\}_{t=1}^T$ . Alternatively, we can understand that  $\text{Var}^b(\cdot)$  is the variance conditional on the observed sample. Note that by choosing an arbitrarily large  $B$ , the feasible bootstrap variance  $\hat{\Gamma}_{n,r}^b$ , defined in (5.1), gives an arbitrarily accurate approximation to the true bootstrap variance  $\text{Var}^b[S(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi})]$ . We have the following proposition.

**Proposition 5.1** *Under Assumptions G1, G2, and R(i)-(iv), if the initial observations  $y_0$  are exogenously given, then  $\frac{1}{nT} [\text{Var}^b(S^r(\hat{Y}_{-1}^b, \hat{u}^b, \hat{\psi})) - \Gamma_{r,n}(\psi_0)] = o_p(1)$ .*

## 5.2 Random effects model with endogenous initial values

When the initial observations  $y_0$  are endogenously given, the disturbance vector now becomes  $(u_0, u_1, u_2, \dots, u_T)$  such that  $\text{Var}(u_0) = \sigma_{v0}^2 \omega_{11}$  and  $\text{Var}(u_t) = \sigma_{v0}^2 \Sigma(\lambda_0, \phi_{\mu 0}), t = 1, \dots, T$ , where  $\omega_{11}$  is defined above (3.31) and  $\Sigma(\lambda, \phi_{\mu})$  is defined in Section 5.1. Define the *transformed residuals*:  $r_0 = \omega_{11}^{-\frac{1}{2}} u_0$ , and  $r_t = \Sigma^{-\frac{1}{2}}(\lambda_0, \phi_{\mu 0}) u_t, t = 1, \dots, T$ , where  $\omega_{11}^{\frac{1}{2}}$  is a square-root matrix of  $\omega_{11}$ . Now, denote the QML estimates of the transformed residuals as  $\{\hat{r}_0, \hat{r}_1, \dots, \hat{r}_T\}$ , and the EDF of the centered  $\hat{r}_t$  by

<sup>12</sup> Although the elements  $\{r_{it}\}$  of  $r_t$  are uncorrelated with constant mean and variance, they may not be totally independent and may not have constant third and fourth moments, unless both  $\mu$  and  $v_t$  are normal.

$\hat{\mathcal{F}}_{n,t}, t = 0, 1, \dots, T$ . Draw  $T + 1$  *matched* samples of size  $n$  each from  $\{\mathcal{F}_{n,t}\}_{t=0}^T$ , to give bootstrap residuals  $\{\hat{r}_0^b, \hat{r}_1^b, \dots, \hat{r}_T^b\}$ . Let  $\hat{\omega}_{11}^{\frac{1}{2}}$  be the plug-in estimator of  $\omega_{11}^{\frac{1}{2}}$ . The bootstrap values for the response variables are thus generated according to

$$\hat{y}_0^b = \tilde{\mathbf{x}}\hat{\pi} + \hat{\omega}_{11}^{\frac{1}{2}}\hat{r}_0^b, \text{ and } \hat{y}_t^b = \hat{\rho}\hat{y}_{t-1}^b + x_t\hat{\beta} + z_t\hat{\gamma} + \Sigma^{\frac{1}{2}}(\hat{\lambda}, \hat{\phi}_\mu)\hat{r}_t^b, t = 1, 2, \dots, T.$$

The rest is analogous to those described in Section 5.1, including the justifications for the validity of this bootstrap procedure. Formally, we have the result for  $S^{rr}(Y_{-1}, u_0, u, \psi_0)$  defined below Theorem 4.3, now written in terms of the lagged response  $Y_{-1}$ , the disturbance vectors  $u_0$  and  $u$ , and the true parameter vector  $\psi_0$ .

**Proposition 5.2** *Under Assumptions G1, G2, R0, and R(i)-(iv\*), if the initial observations  $y_0$  are endogenously given, then  $\frac{1}{n(T+1)}[\text{Var}^b(S^{rr}(\hat{Y}_{-1}^b, \hat{u}_0^b, \hat{u}^b, \hat{\psi})) - \Gamma_{rr,n}(\psi_0)] = o_p(1)$ .*

### 5.3 Fixed effects model with endogenous initial values

When the individual effects are treated as fixed, and the initial differences are modelled by (3.19), the disturbance vector becomes after first-differencing:  $(\Delta\tilde{u}_1, \Delta u_2, \dots, \Delta u_T)$ , where  $\Delta\tilde{u}_1$  is defined in (3.19) and  $\Delta u_t = B_0^{-1}v_t$  as in (3.15) such that  $\text{Var}(\Delta\tilde{u}_1) = \sigma_{v0}^2(\phi_{\zeta 0}I_n + c_m(B_0'B_0)^{-1}) \equiv \sigma_{v0}^2\omega$  and  $\text{Var}(\Delta u_t) = 2\sigma_{v0}^2(B_0'B_0)^{-1}$ ,  $t = 2, \dots, T$ . Define the *transformed residuals*:  $r_1 = \omega^{-\frac{1}{2}}\Delta\tilde{u}_1$  and  $r_t = \frac{1}{\sqrt{2}}B_0\Delta u_t$ ,  $t = 2, \dots, T$ , where  $\omega^{\frac{1}{2}}$  is square-root matrix of  $\omega$ . Denote the QML estimates of the transformed residuals as  $\{\hat{r}_1, \hat{r}_2, \dots, \hat{r}_T\}$ , and the EDF of the centered  $\hat{r}_t$  by  $\hat{\mathcal{F}}_{n,t}$ ,  $t = 1, \dots, T$ . Draw  $T$  *matched* samples of size  $n$  each from  $\{\mathcal{F}_{n,t}\}_{t=1}^T$ , to give bootstrap residuals  $\{\hat{r}_1^b, \hat{r}_2^b, \dots, \hat{r}_T^b\}$ . Let  $\hat{\omega}^{\frac{1}{2}}$  be the plug-in estimator of  $\omega^{\frac{1}{2}}$ . The bootstrap values for the response variables are thus generated according to

$$\Delta\hat{y}_1^b = \Delta x_1\hat{\beta} + \Delta\tilde{\mathbf{x}}\hat{\pi} + \hat{\omega}^{\frac{1}{2}}\hat{r}_1^b, \text{ and } \Delta\hat{y}_t^b = \hat{\rho}\Delta\hat{y}_{t-1}^b + \Delta x_t\hat{\beta} + \sqrt{2}\hat{B}^{-1}\hat{r}_t^b, t = 2, 3, \dots, T.$$

The rest is analogous to those described in Section 5.1, including the justifications for the validity of this bootstrap procedure. Let  $S^f(\Delta Y_{-1}, \Delta u, \psi_0)$  be the score function given below Theorem 4.5, now written in terms of  $\Delta Y_{-1} = \{\Delta y'_t, \dots, \Delta y'_{T-1}\}'$ ,  $\Delta u = \{\Delta\tilde{u}'_1, \Delta u'_2, \dots, \Delta u'_T\}'$ , and  $\psi_0$ . We have the following.

**Proposition 5.3** *Under Assumptions G1, G2, F0, and F(i)-(iv), for either exogenous or endogenous initial observations  $y_0$ , we have  $\frac{1}{nT}[\text{Var}^b(S^f(\Delta\hat{Y}_{-1}^b, \Delta\hat{u}^b, \hat{\psi})) - \Gamma_{f,n}(\psi_0)] = o_p(1)$ .*

## 6 Finite Sample Properties of the QMLEs

Monte Carlo experiments are carried out to investigate the performance of the QMLEs in finite samples and that of the bootstrapped estimates of the standard errors. In the former case, we investigate the consequences of treating the initial observations as endogenous when they are in fact exogenous, and vice versa. In the latter case we study the performance of standard error estimates based on only the Hessian, or only the bootstrapped variance of the score, or both, when errors are normal or nonnormal. We use the following data generating process (DGP):

$$\begin{aligned} y_t &= \rho y_{t-1} + \beta_0 \iota_n + x_t \beta_1 + z_t \gamma + u_t \\ u_t &= \mu + \varepsilon_t \\ \varepsilon_t &= \lambda W_n \varepsilon_t + v_t \end{aligned}$$

where  $y_t, y_{t-1}, x_t$ , and  $z$  are all  $n \times 1$  vectors. The elements of  $x_t$  are generated in a similar fashion as in Hsiao et al. (2002),<sup>13</sup> and the elements of  $z$  are randomly generated from *Bernoulli*(0.5). The spatial weight matrix is generated according to Rook or Queen contiguity, by randomly allocating the  $n$  spatial units on a lattice of  $k \times m$  ( $\geq n$ ) squares, finding the neighbors for each unit, and then row normalizing. We choose  $\beta_0 = 5, \beta_1 = 1, \gamma = 1, \sigma_\mu = 1, \sigma_v = 1$ , a set of values for  $\rho$  ranging from  $-0.9$  to  $0.9$ , a set of values for  $\lambda$  in a similar range,  $T = 3$  or  $7$ , and  $n = 50$  or  $100$ . Each set of Monte Carlo results (corresponding to a combination of the  $\rho$  and  $\lambda$  values) is based on 1000 samples. For bootstrapping standard errors, the number of bootstrap samples is chosen to be  $B = 999 + \lfloor n^{0.75} \rfloor$  where  $\lfloor \cdot \rfloor$  denotes the integer part of  $\cdot$ . Due to space constraints, only a subset of results are reported. The error ( $v_t$ ) distributions can be (i) normal, (ii) normal mixture (10%  $N(0, 4)$  and 90%  $N(0, 1)$ ), or (iii) centered  $\chi^2(5)$  or  $\chi^2(3)$ . For the case of random effects model,  $\mu$  and  $v_t$  are generated from the same distribution.

**Random effects model.** Table 1 reports the Monte Carlo mean and rmse for the random effects model when the data are generated according to either  $m = 0$  or  $m = 6$ , but the model is estimated under  $m = 0, 6$ , and  $200$ . The results show clearly that a correct treatment on the initial values leads to excellent estimation results in general, but a wrong treatment may give totally misleading results.

Some details are as follows. When the true  $m$  value is  $0$ , i.e.,  $y_0$  is exogenous, estimating the model as if  $m = 6$  or  $200$  can give very poor results when  $\rho$  is large. When  $\rho$  is not large or when  $\rho$  is negative (not reported for brevity), the estimates under a wrong  $m$  value improve but are still far from being satisfactory. In contrast, when the true  $m$  value is  $6$  but are treated as either  $0$  or  $200$ , the resulted estimates are in general quite close to the true estimates except for the case of  $m = 0$  under a large and positive  $\rho$ . This shows that the model estimates are not sensitive to the exact choice of  $m$  when  $y_0$  is endogenous and is treated as endogenous. Comparing the results of Table 1a and 1b, we see that non-normality does not deteriorate the results of a wrong treatment of the initial values in terms of mean, but it does in terms of rmse. We note that, when the true  $m$  value is  $0$  but is treated as  $6$  or  $200$ , the poor performance of the estimates when  $\rho$  is large and positive may be attributed to the fact that the quantities  $z_m(\rho)$  and  $a_m(\rho)$ , given below (3.7) and above (3.11), have  $1 - \rho$  as their denominators.

Table 2 reports the standard errors of the estimates based on (1) only the bootstrapped variance of the score (**seSCb**), (2) only the Hessian matrix (**seHS**), and (3) both the bootstrapped variance of the score and the Hessian (**seHSb**). The results show that when errors are normal, all three methods give averaged standard errors very close to the corresponding Monte Carlo SDs; but when errors are non-normal, only the **seHSb** method gives standard errors close to the corresponding Monte Carlo SDs; see in particular the standard errors of  $\phi_\mu$  and  $\sigma_v^2$ . More results corresponding to other choices of the spatial weight matrices, and other values of  $\rho$  and  $\lambda$  are available from the authors upon request.

**Fixed effects model.** The fixed effects  $\mu$  are generated according to either  $\frac{1}{T} \sum_{t=1}^T x_t + e$  or  $e$ , where  $e$  is generated in the same way as  $\mu$  in the random effects model. The reported results correspond to the former. Table 3 reports the Monte Carlo mean and rmse for the fixed effects model when the data are generated according to either  $m = 0$  or  $m = 6$ , but the model is estimated under  $m = 0, 6$ , and  $200$ . The results show again that a correct treatment on the initial values leads to excellent estimation results in general, and that a wrong treatment on the initial values may lead to misleading results though to a much lesser degree as compared with the case of random effects model. The results corresponding to

<sup>13</sup>The detail is:  $x_t = \mu_x + g t 1_n + \zeta_t$ ,  $(1 - \phi_1 L)\zeta_t = \varepsilon_t + \phi_2 \varepsilon_{t-1}$ ,  $\varepsilon_t \sim N(0, \sigma_1^2 I_n)$ ,  $\mu_x = e + \frac{1}{T+m+1} \sum_{t=-m}^T \varepsilon_t$ , and  $e \sim N(0, \sigma_2^2)$ . Let  $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2)$ . Alternatively, the elements of  $x_t$  can be randomly generated from  $N(0, 4)$ .

uncorrelated fixed effects (unreported for brevity) further reveal that whether the individual effects are correlated with the regressors or not does not affect the performance of the fixed-effects QMLEs.

Some details are as follows. When the true  $m$  value is 0, i.e.,  $y_0$  is exogenous, estimates of the model parameters as if  $m = 6$  or 200 can be poor when  $\rho$  is negative and large. When  $\rho$  is not large or when  $\rho$  is positive (not reported for brevity), the estimates under a wrong  $m$  are quite satisfactory. This shows that the model estimates are less sensitive to the treatment on  $y_0$  when it is endogenous. Comparing the results of Table 3a and 3b, we see that non-normality does not deteriorate the results of a wrong treatment of the initial values in terms of mean, but it does in terms of rmse.

Contrary to the case of random effects model, when the true  $m$  value is 0 but is treated as 6 or 200 the estimates of the fixed effects model are poor when  $\rho$  is large but negative. This may be attributed to the quantity  $c_m(\rho)$  defined below (3.21) which has  $1 + \rho$  as its denominator. Comparing the results for the fixed effects model with those for the random effects model, it seems that the fixed effects model is less sensitive to the treatment of the initial values.

Table 4 reports **seSCb**, **seHS**, and **seHSb** along with the Monte Carlo SDs for comparison. The results show that when errors are normal, all three methods give averaged standard errors very close to the corresponding Monte Carlo SDs; but when errors are non-normal, the standard errors of  $\hat{\sigma}_v^2$  from the **seHSb** method are much closer to the corresponding Monte Carlo SDs than those from the other two methods. More results corresponding to other choices of the spatial weight matrices, and other values of  $\rho$  and  $\lambda$  are available from the authors upon request.

## 7 Concluding Remarks

The asymptotic properties of the quasi maximum likelihood estimators of dynamic panel models with spatial errors are studied in detail under the framework that the cross-sectional dimension  $n$  is large and the time dimension  $T$  is fixed, a typical framework for microeconomics data. Both the random effects and fixed effects models are considered, and the assumptions on the initial values and their impact on the subsequent analyses are investigated. The difficulty in implementing the robust standard error estimates, due to the existence of higher order moments of error components in the variance of the score function, is overcome by a simple residual-based bootstrap method. Monte Carlo simulation shows that both the QML estimators and the bootstrap standard errors perform well in finite samples under a correct assumption on initial observations, but the QMLEs can perform poorly when this assumption is not met.

A referee raised a concern that the current paper did not consider the SDPD model with spatial lag dependence, and another referee raised similar concerns on the possible existence of additional spatial structure in the model and on the assumptions made on the processes starting positions. We are fully aware of those intriguing issues. In particular, we recognize that the presence of a spatial lagged dependent variable will complicate the study to a great extent, which certainly demands separate future research.



## Appendix A: Hessian and VC Matrix of Score

**Random effects model with exogenous  $y_0$ .** The Hessian matrix  $H_{rn}(\psi)$  has the elements:

$$\begin{aligned}\frac{\partial^2 \mathcal{L}^r(\psi)}{\partial \theta \partial \theta'} &= -\frac{1}{\sigma_v^2} \tilde{X}' \Omega^{-1} \tilde{X}, & \frac{\partial^2 \mathcal{L}^r(\psi)}{\partial \theta \partial \sigma_v^2} &= -\frac{1}{\sigma_v^4} \tilde{X}' \Omega^{-1} u(\theta), \\ \frac{\partial^2 \mathcal{L}^r(\psi)}{\partial \theta \partial \omega} &= -\frac{1}{\sigma_v^2} \tilde{X}' P_\omega u(\theta), \quad \omega = \lambda, \phi_\mu, & \frac{\partial^2 \mathcal{L}^r(\psi)}{\partial \sigma_v^2 \partial \sigma_v^2} &= -\frac{1}{\sigma_v^6} u(\theta)' \Omega^{-1} u(\theta) + \frac{nT}{2\sigma_v^4}, \\ \frac{\partial^2 \mathcal{L}^r(\psi)}{\partial \sigma_v^2 \partial \omega} &= -\frac{1}{2\sigma_v^4} u(\theta)' P_\omega u(\theta), \quad \omega = \lambda, \phi_\mu, & \frac{\partial^2 \mathcal{L}^r(\psi)}{\partial \omega \partial \omega} &= q_{\omega\omega}[u(\theta)], \quad \text{for } \omega, \varpi = \lambda, \phi_\mu,\end{aligned}$$

where  $q_{\omega\varpi}(u) \equiv \frac{1}{2} \text{tr}(P_\varpi \Omega_\omega - \Omega^{-1} \Omega_{\omega\varpi}) - \frac{1}{2\sigma_v^2} u'(2P_\varpi \Omega_\omega - \Omega^{-1} \Omega_{\omega\varpi}) \Omega^{-1} u$  for  $\omega, \varpi = \lambda, \phi_\mu$ ;  $P_\omega$  is defined below Theorem 4.1; and  $\Omega_{\omega\varpi} = \frac{\partial^2}{\partial \omega \partial \varpi} \Omega(\delta)$  for  $\omega, \varpi = \lambda, \phi_\mu$ . It is easy to see that  $\Omega_{\lambda\lambda} = I_T \otimes \dot{A}$  where  $\dot{A} = \frac{\partial}{\partial \lambda} A = 2(B' B)^{-1}[(W' B + B' W)A - W' W]$ , and all other  $\Omega_{\omega\varpi}$  matrices are  $0_{nT \times nT}$ .

The VC matrix of the score,  $\Gamma_{r,n}(\psi_0) \equiv E[\frac{\partial}{\partial \psi} \mathcal{L}^r(\psi_0) \frac{\partial}{\partial \psi'} \mathcal{L}^r(\psi_0)]$ , has the elements, for  $\omega, \varpi = \lambda, \phi_\mu$ :

$$\begin{aligned}\Gamma_{r,\theta\theta} &= \frac{1}{\sigma_v^2} E(\tilde{X}' \Omega_0^{-1} \tilde{X}), & \Gamma_{r,\theta\sigma_v^2} &= \frac{1}{2\sigma_v^6} E(\tilde{X}' \Omega_0^{-1} u u' \Omega_0^{-1} u), \\ \Gamma_{r,\theta\omega} &= \frac{1}{2\sigma_v^4} E(\tilde{X}' \Omega_0^{-1} u u' P_{\omega 0} u), & \Gamma_{r,\sigma_v^2 \sigma_v^2} &= \frac{1}{\sigma_v^4} g(\Omega_0^{-1}, \Omega_0^{-1}), \\ \Gamma_{r,\sigma_v^2 \omega} &= \frac{1}{\sigma_v^6} g(\Omega_0^{-1}, P_{\omega 0}), & \Gamma_{r,\omega\varpi} &= g(P_{\omega 0}, P_{\varpi 0}),\end{aligned}$$

where  $g(C_1, C_2) \equiv \frac{1}{4\sigma_v^4} E(u' C_1 u u' C_2 u) - \frac{1}{4} \text{tr}(C_1 \Omega_0) \text{tr}(C_2 \Omega_0)$ . The explicit form of  $g$  can be obtained from Lemma B.4(1). The other elements can be evaluated using  $Y_{-1} = \eta_{-1} + (\mathcal{J}'_{\rho_0} \otimes I_n)u$  detail of which can be found in the proof of Proposition 5.1 in Appendix D.

**Random effects model with endogenous  $y_0$ .** The Hessian matrix  $H_{rr,n}(\psi)$  has the elements:

$$\begin{aligned}\frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \theta \partial \theta'} &= -\frac{1}{\sigma_v^2} X^{*'} \Omega^{*-1} X^*, \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \theta \partial \sigma_v^2} &= -\frac{1}{\sigma_v^4} X^{*'} \Omega^{*-1} u^*(\theta, \rho), \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \theta \partial \rho} &= \frac{1}{\sigma_v^2} X_\rho^{*'} \Omega^{*-1} u^*(\theta, \rho) + \frac{1}{\sigma_v^2} X^{*'} \Omega^{*-1} u_\rho^*(\theta, \rho) - \frac{1}{\sigma_v^2} X^{*'} P_\rho^* u^*(\theta, \rho), \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \theta \partial \omega} &= -\frac{1}{\sigma_v^2} X^{*'} P_\omega^* u^*(\theta, \rho), \quad \text{for } \omega = \lambda, \phi_\mu, \text{ and } \phi_\zeta, \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \sigma_v^2 \partial \sigma_v^2} &= -\frac{1}{\sigma_v^6} u^*(\theta, \rho)' \Omega^{*-1} u^*(\theta, \rho) + \frac{n(T+1)}{2\sigma_v^4}, \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \sigma_v^2 \partial \rho} &= \frac{1}{\sigma_v^4} u_\rho^*(\theta, \rho)' \Omega^{*-1} u^*(\theta, \rho) - \frac{1}{2\sigma_v^4} u^*(\theta)' P_\rho^* u^*(\theta, \rho), \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \sigma_v^2 \partial \omega} &= -\frac{1}{2\sigma_v^4} u^*(\theta, \rho)' P_\omega^* u^*(\theta, \rho), \quad \text{for } \omega = \lambda, \phi_\mu, \text{ and } \phi_\zeta, \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \rho \partial \rho} &= -\frac{1}{\sigma_v^2} u_\rho^*(\theta, \rho)' \Omega^{*-1} u_\rho^*(\theta, \rho) - \frac{1}{\sigma_v^2} u_\rho^*(\theta, \rho)' \Omega^{*-1} u_\rho^*(\theta, \rho) + \frac{2}{\sigma_v^2} u_\rho^*(\theta, \rho)' P_\rho^* u^*(\theta, \rho) + q_{\rho\rho}^*[u^*(\theta, \rho)], \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \rho \partial \omega} &= \frac{1}{\sigma_v^2} u_\rho^*(\theta, \rho)' P_\omega^* u^*(\theta, \rho) + q_{\rho\omega}^*[u^*(\theta, \rho)], \quad \text{for } \omega = \lambda, \phi_\mu, \text{ and } \phi_\zeta, \\ \frac{\partial^2 \mathcal{L}^{rr}(\psi)}{\partial \omega \partial \omega} &= q_{\omega\omega}^*[u^*(\theta, \rho)], \quad \text{for } \omega, \varpi = \lambda, \phi_\mu, \text{ and } \phi_\zeta.\end{aligned}$$

where  $q_{\omega\varpi}^*(u^*) \equiv \frac{1}{2} \text{tr}(P_\varpi^* \Omega_\omega^* - \Omega^{*-1} \Omega_{\omega\varpi}^*) - \frac{1}{2\sigma_v^2} u^{*'}(2P_\varpi^* \Omega_\omega^* - \Omega^{*-1} \Omega_{\omega\varpi}^*) \Omega^{*-1} u^*$  for  $\omega, \varpi = \rho, \lambda, \phi_\mu$ , and  $\phi_\zeta$ ,  $X_\rho^* = \frac{\partial}{\partial \rho} X^*$ ,  $u_{\rho\rho}^*(\theta, \rho) = \frac{\partial^2}{\partial \rho^2} u^*(\theta, \rho)$ , and  $\Omega_{\rho\omega}^* = \frac{\partial^2}{\partial \rho \partial \omega} \Omega^*$  for  $\omega = \rho, \lambda, \phi_\mu$ , and  $\phi_\zeta$ . The second-order partial derivatives of  $\Omega^*$  are

$$\begin{aligned}\Omega_{\rho\rho}^* &= \begin{pmatrix} 2\phi_\mu(\dot{a}_m^2 + \ddot{a}_m)I_n + \ddot{b}_m(B' B)^{-1}, & \phi_\mu \ddot{a}_m(\iota' \otimes I_n) \\ \phi_\mu \ddot{a}_m(\iota \otimes I_n) & 0_{nT \times nT} \end{pmatrix}, \quad \Omega_{\rho\lambda}^* = \begin{pmatrix} \dot{b}_m A, & 0_{n \times nT} \\ 0_{nT \times n} & 0_{nT \times nT} \end{pmatrix}, \\ \Omega_{\rho\phi_\mu}^* &= \begin{pmatrix} 2a_m \dot{a}_m I_n, & \dot{a}_m(\iota' \otimes I_n) \\ \dot{a}_m(\iota \otimes I_n) & 0_{nT \times nT} \end{pmatrix}, \quad \Omega_{\lambda\lambda}^* = \begin{pmatrix} b_m & 0 \\ 0 & I_T \end{pmatrix} \otimes \dot{A},\end{aligned}$$

and all other  $\Omega_{\omega\varpi}^*$  matrices are  $0_{n(T+1) \times n(T+1)}$ , where  $\ddot{a}_m = \frac{\partial}{\partial \rho} \dot{a}_m$  and  $\ddot{b} = \frac{\partial}{\partial \rho} \dot{b}_m$  and their exact expressions can be easily derived. Finally,  $X_\rho^*$  has a sole non-zero element  $\dot{a}_m z$ , and  $u_{\rho\rho}^*(\theta, \rho) = (-\ddot{a}_m \gamma' z', 0_{1 \times nT})'$ .

The VC matrix of the score,  $\Gamma_{rr,n}(\psi_0) \equiv E[\frac{\partial}{\partial \psi} \mathcal{L}^{rr}(\psi_0) \frac{\partial}{\partial \psi'} \mathcal{L}^{rr}(\psi_0)]$ , has the elements, for  $\omega$  and  $\varpi = \lambda, \phi_\mu$ , or  $\phi_\zeta$ :

$$\begin{aligned} \Gamma_{rr,\theta\theta} &= \frac{1}{\sigma_{v0}^2} E(X^{*'} \Omega_0^{*-1} X^*), & \Gamma_{rr,\theta\sigma_v^2} &= \frac{1}{\sigma_{v0}^2} f_1^*(\Omega_0^{*-1}), \\ \Gamma_{rr,\theta\rho} &= f_1^*(P_{\rho0}^*) - f_2^*(\Omega_0^{*-1}), & \Gamma_{rr,\theta\omega} &= g_1^*(P_{\omega0}^*), \\ \Gamma_{rr,\sigma_v^2\sigma_v^2} &= \frac{1}{\sigma_{v0}^4} g_1^*(\Omega_0^{*-1}, \Omega_0^{*-1}), & \Gamma_{rr,\sigma_v^2\rho} &= \frac{1}{\sigma_{v0}^2} [g_1^*(P_{\rho0}^*, \Omega_0^{*-1}) - g_2^*(\Omega_0^{*-1}, \Omega_0^{*-1})], \\ \Gamma_{rr,\sigma_v^2\omega} &= \frac{1}{\sigma_{v0}^2} g_1^*(\Omega_0^{*-1}, P_{\omega0}^*), & \Gamma_{rr,\rho\rho} &= \frac{1}{\sigma_{v0}^4} E[(u_\rho^{*'} \Omega_0^{*-1} u_\rho^*)^2] + g_1^*(P_{\rho0}^*, P_{\rho0}^*) - 2g_2^*(\Omega_0^{*-1}, P_{\rho0}^*), \\ \Gamma_{rr,\rho\omega} &= g_1^*(P_{\rho0}^*, P_{\omega0}^*) - g_2^*(\Omega_0^{*-1}, P_{\omega0}^*), & \Gamma_{rr,\omega\omega} &= g_1^*(P_{\omega0}^*, P_{\omega0}^*), \end{aligned}$$

where  $f_1^*(A) \equiv \frac{1}{2\sigma_{v0}^4} E(X^{*'} \Omega_0^{*-1} u^* u^{*'} A u^*)$ ,  $f_2^*(A) \equiv \frac{1}{\sigma_{v0}^4} E(X^{*'} \Omega_0^{*-1} u^* u_\rho^{*'} A u^*)$ ,  $P_\omega^*$  is defined below Theorem 4.3,  $g_1^*(A, B) \equiv \frac{1}{4\sigma_{v0}^4} E(u^{*'} A u^* u^{*'} B u^*) - \frac{1}{4} \text{tr}(A \Omega_0^*) \text{tr}(B \Omega_0^*)$ , and  $g_2^*(A, B) \equiv \frac{1}{4\sigma_{v0}^4} E(u_\rho^{*'} A u^* u^{*'} B u^*)$ . As  $X^*$  is exogenous, the explicit forms of  $f_1^*$  and  $g_1^*$  can be obtained from Lemma B.4. The functions  $f_2^*$  and  $g_2^*$  can be evaluated using  $u_\rho^* = -(\dot{a}_m \gamma' Z', Y_{-1}')' = -\eta_{-1}^* - (\mathcal{J}_\rho^{*'} \otimes I_n) u^*$  given in the proof of Proposition 5.2 in Appendix D.

**Fixed effects model with exogenous or endogenous  $y_0$ .** The Hessian matrix  $H_{f,n}(\psi)$  has the elements:

$$\begin{aligned} \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \theta \partial \theta'} &= -\frac{1}{\sigma_v^2} \Delta X^{\dagger'} \Omega^{\dagger-1} \Delta X^\dagger, \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \theta \partial \sigma_v^2} &= -\frac{1}{\sigma_v^4} \Delta X^{\dagger'} \Omega^{\dagger-1} \Delta u(\theta, \rho), \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \theta \partial \rho} &= \frac{1}{\sigma_v^2} \Delta X^{\dagger'} \Omega^{\dagger-1} \Delta u_\rho(\theta, \rho) - \frac{1}{\sigma_v^2} \Delta X^{\dagger'} P_\rho^\dagger \Delta u(\theta, \rho), \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \theta \partial \omega} &= -\frac{1}{\sigma_v^2} \Delta X^{\dagger'} P_\omega^\dagger \Delta u(\theta, \rho), \text{ for } \omega = \lambda, \phi_\zeta, \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \sigma_v^2 \partial \sigma_v^2} &= -\frac{1}{\sigma_v^6} \Delta u(\theta, \rho)' \Omega^{\dagger-1} \Delta u(\theta, \rho) + \frac{nT}{2\sigma_v^4}, \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \sigma_v^2 \partial \rho} &= \frac{1}{\sigma_v^4} \Delta u_\rho(\theta, \rho)' \Omega^{\dagger-1} \Delta u(\theta, \rho) - \frac{1}{2\sigma_v^4} \Delta u(\theta, \rho)' P_\rho^\dagger \Delta u(\theta, \rho), \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \sigma_v^2 \partial \omega} &= -\frac{1}{2\sigma_v^4} \Delta u(\theta, \rho)' P_\omega^\dagger \Delta u(\theta, \rho), \text{ for } \omega = \lambda, \phi_\zeta, \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \rho \partial \rho} &= -\frac{1}{\sigma_v^2} \Delta u_\rho(\theta, \rho)' \Omega^{\dagger-1} \Delta u_\rho(\theta, \rho) + \frac{2}{\sigma_v^2} \Delta u_\rho(\theta, \rho)' P_\rho^\dagger \Delta u(\theta, \rho) + q_{\rho\rho}^\dagger[\Delta u(\theta, \rho)], \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \rho \partial \omega} &= \frac{1}{\sigma_v^2} \Delta u_\rho(\theta, \rho)' P_\omega^\dagger \Delta u(\theta, \rho) + q_{\rho\omega}^\dagger[\Delta u(\theta, \rho)], \text{ for } \omega = \lambda, \phi_\zeta, \\ \frac{\partial^2 \mathcal{L}^f(\psi)}{\partial \omega \partial \omega} &= q_{\omega\omega}^\dagger[\Delta u(\theta, \rho)], \text{ for } \omega, \varpi = \lambda, \phi_\zeta, \end{aligned}$$

where  $q_{\omega\varpi}^\dagger(\Delta u) \equiv \frac{1}{2} \text{tr}(P_\omega^\dagger \Omega_\omega^\dagger - \Omega^{\dagger-1} \Omega_{\omega\varpi}^\dagger) - \frac{1}{2\sigma_v^2} \Delta u'(2P_\omega^\dagger \Omega_\omega^\dagger - \Omega^{\dagger-1} \Omega_{\omega\varpi}^\dagger) \Omega^{\dagger-1} \Delta u$  for  $\omega, \varpi = \rho, \lambda$ , and  $\phi_\zeta$ . The second derivatives  $\Omega_{\omega\varpi}$  of  $\Omega$  are:  $\Omega_{\rho\rho} = h_{\ddot{c}_m} \otimes (B'B)^{-1}$  where  $\ddot{c}_m = \frac{\partial}{\partial \rho} \dot{c}_m$ ,  $\Omega_{\rho\lambda} = h_{\dot{c}_m} \otimes A$ ,  $\Omega_{\lambda\lambda} = h_{c_m} \otimes \dot{A}$ , and the remaining are all zero matrices.

The VC matrix of the score,  $\Gamma_{f,n}(\psi_0) = E[\frac{\partial}{\partial \psi} \mathcal{L}^f(\psi_0) \frac{\partial}{\partial \psi'} \mathcal{L}^f(\psi_0)]$ , has the elements, for  $\omega, \varpi = \lambda, \phi_\zeta$ :

$$\begin{aligned} \Gamma_{f,\theta\theta} &= \frac{1}{\sigma_{v0}^2} E(\Delta X^{\dagger'} \Omega_0^{\dagger-1} \Delta X^\dagger), & \Gamma_{f,\theta\sigma_v^2} &= \frac{1}{\sigma_{v0}^2} f_1^\dagger(\Omega_0^{\dagger-1}), \\ \Gamma_{f,\theta\rho} &= f_1^\dagger(P_{\rho0}^\dagger) - f_2^\dagger(\Omega_0^{\dagger-1}), & \Gamma_{f,\theta\omega} &= f_1^\dagger(P_{\omega0}^\dagger), \\ \Gamma_{f,\sigma_v^2\sigma_v^2} &= \frac{1}{\sigma_{v0}^4} g_1^\dagger(\Omega_0^{\dagger-1}, \Omega_0^{\dagger-1}), & \Gamma_{f,\sigma_v^2\rho} &= \frac{1}{\sigma_{v0}^2} [g_1^\dagger(P_{\rho0}^\dagger, \Omega_0^{\dagger-1}) - g_2^\dagger(\Omega_0^{\dagger-1}, \Omega_0^{\dagger-1})], \\ \Gamma_{f,\sigma_v^2\omega} &= \frac{1}{\sigma_{v0}^2} g_1^\dagger(\Omega_0^{\dagger-1}, P_{\omega0}^\dagger), & \Gamma_{f,\rho\rho} &= \frac{1}{\sigma_{v0}^4} E[(\Delta u_\rho^{\dagger'} \Omega_0^{\dagger-1} \Delta u_\rho^\dagger)^2] + g_1^\dagger(P_{\rho0}^\dagger, P_{\rho0}^\dagger) - 2g_2^\dagger(\Omega_0^{\dagger-1}, P_{\rho0}^\dagger), \\ \Gamma_{f,\rho\omega} &= g_1^\dagger(P_{\rho0}^\dagger, P_{\omega0}^\dagger) - g_2^\dagger(\Omega_0^{\dagger-1}, P_{\omega0}^\dagger), & \Gamma_{f,\omega\omega} &= g_1^\dagger(P_{\omega0}^\dagger, P_{\omega0}^\dagger), \end{aligned}$$

where  $f_1^\dagger(A) \equiv \frac{1}{2\sigma_{v0}^4} E(\Delta X^{\dagger'} \Omega_0^{\dagger-1} \Delta u \Delta u' A \Delta u)$ ,  $f_2^\dagger(A) \equiv \frac{1}{\sigma_{v0}^4} E(\Delta X^{\dagger'} \Omega_0^{\dagger-1} \Delta u \Delta u_\rho' A \Delta u)$ ,  $g_1^\dagger(A, B) \equiv \frac{1}{4\sigma_{v0}^4} E(\Delta u' A \Delta u \Delta u' B \Delta u) - \frac{1}{4} \text{tr}(A \Omega_0^\dagger) \text{tr}(B \Omega_0^\dagger)$ ,  $g_2^\dagger(A, B) \equiv \frac{1}{4\sigma_{v0}^4} E(\Delta u_\rho' A \Delta u \Delta u_\rho' B \Delta u)$ , and  $P_\omega^\dagger$  is defined below Theorem 4.5. As  $\Delta X^\dagger$  is exogenous, the explicit forms of  $f_1^\dagger$  and  $g_1^\dagger$  can be obtained from Lemma B.4. The functions  $f_2^\dagger$  and  $g_2^\dagger$  can be evaluated using  $\Delta u_\rho = -(\mathcal{O}'_{n \times 1}, \Delta y'_1, \dots, \Delta y'_{T-1})' = -\Delta \eta_{-1} - (\mathcal{J}_\rho \otimes I_n) \Delta u$  given in the proof of Proposition 5.3 in Appendix D.

## Appendix B: Some Useful Lemmas

We introduce some technical lemmas that are used in the proofs of the main results. The proofs of all lemmas are provided in a supplementary material that is made available online at [http://www.mysmu.edu/faculty/ljsu/Publications/Panel\\_qmle\\_supp.pdf](http://www.mysmu.edu/faculty/ljsu/Publications/Panel_qmle_supp.pdf).

We first state five lemmas that greatly facilitate the proof of subsequent lemmas and some results in the main theorems.

**Lemma B.1** *Let  $P_n$  and  $Q_n$  be two  $n \times n$  matrices that are uniformly bounded in both row and column sums. Let  $R_n$  be a conformable matrix whose elements are uniformly  $O(o_n)$  for a certain sequence  $o_n$ . Then we have:*

- (1)  $P_n Q_n$  is also uniformly bounded in both row and column sums;
- (2) any  $(i, j)$  elements  $P_{n,ij}$  of  $P_n$  are uniformly bounded in  $i$  and  $j$  and  $\text{tr}(P_n) = O(n)$ ;
- (3) the elements of  $P_n R_n$  and  $R_n P_n$  are uniformly  $O(o_n)$ .

**Lemma B.2** *Suppose that Assumption G2 holds.*

- (1)  $B'B, (B'B)^{-1}, \Omega, \Omega^*, \Omega^\dagger, A$ , and  $\dot{A}$  are all uniformly bounded in both row and column sums.
- (2)  $\frac{1}{n} \text{tr}(D_1 \Omega D_2) = O(1)$  for  $D_1, D_2 = \Omega^{-1}, \Omega^{-1}(I_T \otimes A)\Omega^{-1}, \Omega^{-1}(J_T \otimes I_n)\Omega^{-1}$ , and  $\Omega^{-1}(I_T \otimes \dot{A})$ . The same conclusion holds when  $\Omega$  is replaced by  $\Omega^*$  or  $\Omega^\dagger$ , and  $D_1$  and  $D_2$  are replaced by their analogs corresponding to the case of  $\Omega^*$  or  $\Omega^\dagger$ .
- (3)  $\frac{1}{n} \text{tr}(B'^{-1} R B^{-1}) = O(1)$  where  $R$  is an  $n \times n$  nonstochastic matrix that is uniformly bounded in both row and column sums.

**Lemma B.3** *Let  $\{a_i\}_{i=1}^n$  and  $\{b_i\}_{i=1}^n$  be two independent iid sequences with zero means and fourth moments. Let  $\sigma_a^2 = E(a_1^2), \sigma_b^2 = E(b_1^2)$ . Let  $q_n$  and  $p_n$  be  $n \times n$  nonstochastic matrices. Then*

- (1)  $E[(a' q_n a)(a' p_n a)] = \kappa_a \sum_{i=1}^n q_{n,ii} p_{n,ii} + \sigma_a^4 [\text{tr}(q_n) \text{tr}(p_n) + \text{tr}(q_n(p_n + p'_n))]$ ,
- (2)  $E[(a' q_n a)(b' p_n b)] = \sigma_a^2 \sigma_b^2 \text{tr}(q_n) \text{tr}(p_n)$ ,
- (3)  $E[(a' q_n b)(a' p_n b)] = \sigma_a^2 \sigma_b^2 \text{tr}(q_n p'_n)$ ,

where  $\kappa_a \equiv E(a^4) - 3[E(a^2)]^2$ , and, e.g.,  $q_{n,ij}$  denotes the  $(i, j)$ th element of  $q_n$ .

**Lemma B.4** *Recall  $u = (\iota_T \otimes I_n)\mu + (I_T \otimes B_0^{-1})v$ . Let  $a = \zeta + \mu(1 - \rho_0^m)/(1 - \rho_0) + \sum_{j=0}^{m-1} \rho_0^j B_0^{-1} v_{-j}$ , where  $\zeta, \mu$ , and  $v$  are defined in the text. In particular,  $\zeta_i$ 's are iid and independent of  $\mu$  and  $v$ . Let  $q_n, p_n, r_n, s_n, t_n$  be  $nT \times nT, nT \times nT, n \times n, n \times nT$  and  $n \times nT$  nonstochastic matrices, respectively. Further,  $q_n, p_n$ , and  $r_n$  are symmetric. Then*

- (1)  $E[(u' q_n u)(u' p_n u)] = \kappa_\mu \sum_{i=1}^n G_{q_n,1ii} G_{p_n,1ii} + \kappa_v \sum_{i=1}^{nT} G_{q_n,2ii} G_{p_n,2ii} + \sigma_v^4 [\text{tr}(q_n \Omega_0) \text{tr}(p_n \Omega_0) + 2 \text{tr}(q_n \Omega_0 p_n \Omega_0)]$ ,
- (2)  $E[(u' q_n u)(a' r_n a)] = \frac{\kappa_\mu (1 - \rho_0^m)^2}{(1 - \rho_0)^2} \sum_{i=1}^n G_{q_n,1ii} r_{n,ii} + \sigma_v^4 [\text{tr}(r_n \omega_{11}) \text{tr}(q_n \Omega_0) + 2 \text{tr}(\omega_{12} q_n \omega_{21} p_n)]$ ,
- (3)  $E[(a' s_n u)(a' t_n u)] = \frac{\kappa_\mu (1 - \rho_0^m)^2}{(1 - \rho_0)^2} \sum_{i=1}^n (s_n (\iota_T \otimes I_n))_{ii} (t_n (\iota_T \otimes I_n))_{ii} + \sigma_v^4 [\text{tr}(s_n \omega_{21}) \text{tr}(t_n \omega_{21}) + \text{tr}(s_n \omega_{21} t_n \omega_{21}) + \text{tr}(s_n \Omega_0 t'_n \omega_{11})]$ ,
- (4)  $E[(u' q_n u)(u' s'_n a)] = \frac{\kappa_\mu (1 - \rho_0^m)}{1 - \rho_0} \sum_{i=1}^n G_{q_n,1ii} ((\iota'_T \otimes I_n) s'_n)_{ii} + \sigma_v^4 [\text{tr}(q_n \Omega_0) \text{tr}(s'_n \omega_{12}) + 2 \text{tr}(\Omega_0 s'_n \omega_{12} q_n)]$ ,
- (5)  $E[(a' r_n a)(a' s_n u)] = \frac{\kappa_\mu (1 - \rho_0^m)^3}{(1 - \rho_0)^3} \sum_{i=1}^n r_{n,ii} (s_n (\iota_T \otimes I_n))_{ii} + \sigma_v^4 [(r_n \omega_{11}) \text{tr}(s_n \omega_{21}) + 2 \text{tr}(r_n \omega_{11} s_n \omega_{21})]$ ,

where  $G_{q_n,1} \equiv (\iota'_T \otimes I_n) q_n (\iota_T \otimes I_n)$ ,  $G_{q_n,2} \equiv (I_T \otimes B_0^{-1}) q_n (I_T \otimes B_0^{-1})$ , and, e.g.,  $G_{q_n,1ij}$  denotes the  $(i, j)$ th element of  $G_{q_n,1}$ .

**Lemma B.5** Suppose that  $\{P_{1n}\}$  and  $\{P_{2n}\}$  are sequences of matrices with row and column sums uniformly bounded. Let  $a = (a_1, \dots, a_n)'$ , where  $a_i$ 's are independent random variables such that  $\sup_i E|a_i|^{2+\epsilon_0} < \infty$  for some  $\epsilon_0 > 0$ . Let  $b = (b_1, \dots, b_n)'$ , where  $b_i$ 's are iid with mean zero and  $(4 + 2\epsilon_0)$ th finite moments, and  $\{b_i\}$  is independent of  $\{a_i\}$ . Let  $\sigma_{Q_n}^2$  be the variance of  $Q_n = a'P_{1n}b + b'P_{2n}b - \sigma_b^2 \text{tr}(P_{2n})$ . Assume that the elements of  $P_{1n}$ ,  $P_{2n}$  are of uniform order  $O(1/\sqrt{h_n})$  and  $O(1/h_n)$ , respectively. If  $\lim_{n \rightarrow \infty} h_n^{1+2/\epsilon_0}/n = 0$ , then  $Q_n/\sigma_{Q_n} \xrightarrow{d} N(0, 1)$ .

Now, for ease of exposition we assume that both  $x_{it}$  and  $z_i$  are scalar random variables ( $p = 1, q = 1$ ) in this Appendix. For the proofs of Theorems 2 and 4 for the SDPD model with random effects, the following presentations are essential. By continuous backward substitutions, we have for  $t = 0, 1, 2, \dots$ ,

$$y_t = \mathbb{X}_t \beta_0 + c_{\rho_0, t} z \gamma_0 + c_{\rho_0, t} \mu + \mathbb{V}_t + \mathbb{Y}_{0, t}, \quad (\text{B.1})$$

where for fixed  $y_0$ ,  $\mathbb{X}_t = \sum_{j=0}^{t-1} \rho_0^j x_{t-j}$ ,  $\mathbb{V}_t = \sum_{j=0}^{t-1} \rho_0^j B_0^{-1} v_{t-j}$ ,  $\mathbb{Y}_{0, t} = \rho_0^t y_0$  and  $c_{\rho, t} = (1 - \rho^t)/(1 - \rho)$ ; and for endogenous  $y_0$ ,  $\mathbb{X}_t = \sum_{j=0}^{t+m-1} \rho_0^j x_{t-j}$ ,  $\mathbb{V}_t = \sum_{j=0}^{t+m-1} \rho_0^j B_0^{-1} v_{t-j}$ ,  $\mathbb{Y}_{0, t} = \rho_0^{t+m} y_{-m}$ , and  $c_{\rho, t} = (1 - \rho^{t+m})/(1 - \rho)$ . Now, define  $\mathbb{Y}_0 = (\mathbb{Y}'_{0,0}, \mathbb{Y}'_{0,1}, \dots, \mathbb{Y}'_{0,T-1})'$ . Then

$$Y_{-1} = \mathbb{X}_{(-1)} \beta_0 + (l_{\rho_0} \otimes I_n) z \gamma_0 + (l_{\rho_0} \otimes I_n) \mu + \mathbb{V}_{(-1)} + \mathbb{Y}_0, \quad (\text{B.2})$$

where  $\mathbb{X}_{(-1)} = (0, \mathbb{X}'_1, \dots, \mathbb{X}'_{T-1})'$ ,  $\mathbb{V}_{(-1)} = (0, \mathbb{V}'_1, \dots, \mathbb{V}'_{T-1})'$ , and  $l_{\rho} = (0, c_{\rho,1}, \dots, c_{\rho,T-1})'$  when  $y_0$  is fixed, and  $\mathbb{X}_{(-1)} = (\mathbb{X}'_0, \mathbb{X}'_1, \dots, \mathbb{X}'_{T-1})'$ ,  $\mathbb{V}_{(-1)} = (\mathbb{V}'_0, \mathbb{V}'_1, \dots, \mathbb{V}'_{T-1})'$ , and  $l_{\rho} = (c_{\rho,0}, c_{\rho,1}, \dots, c_{\rho,T-1})'$  when  $y_0$  is endogenous. Notice that when  $y_0$  is exogenous,  $Y_{-1}$  can also be expressed as

$$Y_{-1} = A_x X \beta_0 + (l_{\rho_0} \otimes I_n) z \gamma_0 + (l_{\rho_0} \otimes I_n) \mu + A_v v + \mathbb{Y}_0, \quad (\text{B.3})$$

where  $A_x = \mathcal{J}'_{\rho_0} \otimes I_n$  and  $A_v = \mathcal{J}'_{\rho_0} \otimes B_0^{-1}$  with

$$\mathcal{J}_{\rho} = \begin{pmatrix} 0 & 1 & \rho & \cdots & \rho^{T-2} \\ 0 & 0 & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (\text{B.4})$$

Lemmas B.6-B.8 given below are used in the proof of Theorem 4.2.

**Lemma B.6** Under the assumptions of Theorem 4.2,  $E(\tilde{X}' \Omega_0^{-1} u) = 0$ .

**Lemma B.7** Under the assumptions of Theorem 4.2,  $\frac{1}{nT} \left\{ \frac{\partial \mathcal{L}^r(\psi_0)}{\partial \psi \partial \psi'} - E \left[ \frac{\partial \mathcal{L}^r(\psi_0)}{\partial \psi \partial \psi'} \right] \right\} = o_p(1)$ .

**Lemma B.8** Under the assumptions of Theorem of 4.2,  $\frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}^r(\psi_0)}{\partial \psi} \xrightarrow{d} N(0, \Gamma_r)$ .

Lemmas B.9-B.13 are used in the proof of Theorem 4.4, for the SDPD model with random effects and endogenous  $y_0$ . Let  $R_{ts}$  be an  $n \times n$  symmetric and positive semidefinite (p.s.d.) nonstochastic square matrix for  $t, s = 0, 1, \dots, T-1$ . Assume that  $R_{ts}$  are uniformly bounded in both row and column sums. Recall for this case,  $\mathbb{X}_t = \sum_{j=0}^{t+m-1} \rho_0^j x_{t-j}$  and  $\mathbb{V}_t = \sum_{j=0}^{t+m-1} \rho_0^j B_0^{-1} v_{t-j}$ .

**Lemma B.9** Suppose that the conditions in Theorem 4.4 are satisfied. Then

- (1)  $E(\mathbb{V}'_t R_{ts} \mathbb{V}_s) = \sigma_v^2 \text{tr}(B_0'^{-1} R_{ts} B_0^{-1}) \sum_{i=\max(0, t-s)}^{t+m-1} \rho_0^{s-t+2i},$
- (2)  $E(\mathbb{X}'_t R_{ts} \mathbb{X}_s) = \text{tr}(\sum_{j=0}^{s+m-1} \sum_{k=0}^{t+m-1} \rho_0^{j+k} R_{ts} E(x_{s-j} x'_{t-k})),$
- (3)  $E(\mathbb{X}'_t R_{ts} \mathbb{V}_s) = 0.$

**Lemma B.10** Suppose that the conditions in Theorem 4.4 are satisfied. Then

- (1)  $\text{Cov}(\mathbb{V}'_t R_{ts} \mathbb{V}_s, \mathbb{V}'_g R_{gh} \mathbb{V}_h) = \rho_{tsgh,1} \{ \kappa_v \sum_{i=1}^n \bar{B}_{ts,ii} \bar{B}_{gh,ii} + 2\sigma_v^4 \text{tr}[\bar{B}_{ts}(\bar{B}_{gh} + \bar{B}'_{gh})] \}$   
 $+ \rho_{tsgh,2} \sigma_v^4 \text{tr}[B_0'^{-1} R_{ts} (B'_0 B_0)^{-1} R_{gh} B_0^{-1}]$   
 $+ \rho_{tsgh,3} \sigma_v^4 \text{tr}[B_0'^{-1} R_{ts} (B'_0 B_0)^{-1} R'_{gh} B_0^{-1}],$
- (2)  $\text{Cov}(\mathbb{X}'_t R_{ts} \mathbb{V}_s, \mathbb{X}'_g R_{gh} \mathbb{V}_h) = \sigma_v^2 \text{tr}[\sum_{i=0}^{t+m-1} \sum_{k=0}^{g+m-1} \sum_{j=\max(0, s-h)}^{s+m-1} \rho_0^{i+k+h-s+2j} R_{ts}$   
 $\times (B'_0 B_0)^{-1} R'_{gh} E(x'_{g-k} x_{t-i})],$
- (3)  $\text{Cov}(\mathbb{X}'_t R_{ts} \mathbb{X}_s, \mathbb{X}'_g R_{gh} \mathbb{X}_h) = O(n),$

where  $\bar{B}_{ts,ii}$  denotes the  $(i, i)$ th element of  $\bar{B}_{ts} \equiv B_0'^{-1} R_{ts} B_0^{-1}$ ,  $\rho_{tsgh,1} = \sum_{j=\max(0, t-s, t-g, t-h)}^{t+m-1} \rho_0^{(s+g+h-3t+4j)}$ ,  
 $\rho_{tsgh,2} = \sum_{i=\max(0, t-g)}^{t+m-1} \rho_0^{g-t+2i} \sum_{j=\max(0, s-h)}^{s+m-1} \rho_0^{h-s+2j} 1(j \neq i+s-t)$ , and  $\rho_{tsgh,3} = \sum_{i=\max(0, t-h)}^{t+m-1} \rho_0^{h-t+2i}$   
 $\sum_{j=\max(0, s-g)}^{s+m-1} \rho_0^{g-s+2j} 1(j \neq i+s-t).$

**Lemma B.11** Suppose that the conditions in Theorem 4.4 are satisfied. Then

- (1)  $\frac{1}{nT} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} [\mathbb{V}'_t R_{ts} \mathbb{V}_s - E(\mathbb{V}'_t R_{ts} \mathbb{V}_s)] \xrightarrow{p} 0,$
- (2)  $\frac{1}{nT} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \mathbb{X}'_t R_{ts} \mathbb{V}_s \xrightarrow{p} 0,$
- (3)  $\frac{1}{nT} \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} [\mathbb{X}'_t R_{ts} \mathbb{X}_s - E(\mathbb{X}'_t R_{ts} \mathbb{X}_s)] \xrightarrow{p} 0.$

**Lemma B.12** Under the assumptions of Theorem 4.4,  $\frac{1}{n(T+1)} \left\{ \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \psi \partial \psi'} - E \left[ \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \psi \partial \psi'} \right] \right\} = o_p(1).$

**Lemma B.13** Under the assumptions of Theorem 4.4,  $\frac{1}{\sqrt{n(T+1)}} \frac{\partial \mathcal{L}^{rr}(\psi_0)}{\partial \psi} \xrightarrow{d} N(0, \Gamma_{rr}).$

Lemmas B.14-B.15 are used in the proof of Theorem 4.6 for the fixed effects model.

**Lemma B.14** Under the assumptions of Theorem 4.6,  $\frac{1}{nT} \left\{ \frac{\partial \mathcal{L}^f(\psi_0)}{\partial \psi \partial \psi'} - E \left[ \frac{\partial \mathcal{L}^f(\psi_0)}{\partial \psi \partial \psi'} \right] \right\} = o_p(1).$

**Lemma B.15** Suppose that the conditions in Theorem 4.6 are satisfied. Then  $\frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}^f(\psi_0)}{\partial \psi} \xrightarrow{d} N(0, \Gamma_f).$

## Appendix C: Proofs of the Theorems in Section 4

**Proof of Theorem 4.1.** By Theorem 3.4 of White (1994), it suffices to show that: (i)  $\frac{1}{nT} [\mathcal{L}_c^{r*}(\delta) - \mathcal{L}_c^r(\delta)] \xrightarrow{p} 0$  uniformly in  $\delta \in \Delta$ , and (ii)  $\limsup_{n \rightarrow \infty} \max_{\delta \in N_\epsilon^c(\delta_0)} \frac{1}{nT} [\mathcal{L}_c^{r*}(\delta) - \mathcal{L}_c^{r*}(\delta_0)] < 0$  for any  $\epsilon > 0$ , where  $N_\epsilon^c(\delta_0)$  is the complement of an open neighborhood of  $\delta_0$  on  $\Delta$  of radius  $\epsilon$ . By (3.5) and (4.3),  $\frac{2}{nT} [\mathcal{L}_c^{r*}(\delta) - \mathcal{L}_c^r(\delta)] = -\ln \tilde{\sigma}_v^2(\delta) + \ln \hat{\sigma}_v^2(\delta)$ . To show (i), it is sufficient to show

$$\hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta) = o_p(1) \text{ uniformly in } \delta \in \Delta \quad (\text{C.1})$$

and  $\tilde{\sigma}_v^2(\delta)$  is uniformly bounded away from zero on  $\Delta$ . The latter will be checked in the proof of (ii). So we focus on the proof of (C.1) here. By the definition of  $\tilde{u}(\delta)$  below (3.4), we have  $\tilde{u}(\delta) =$

$Y - \tilde{X}(\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1}Y = \Omega^{1/2}M\Omega^{-1/2}Y$  where  $M = I_{nT} - \Omega^{-1/2}\tilde{X}(\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1/2}$  is a projection matrix. This, in conjunction with the fact that  $M\Omega^{-1/2}\tilde{X} = 0$ , implies that

$$\hat{\sigma}_v^2(\delta) = \frac{1}{nT}\tilde{u}(\delta)'\Omega^{-1}\tilde{u}(\delta) = \frac{1}{nT}Y'\Omega^{-1/2}M\Omega^{-1/2}Y = \frac{1}{nT}u'\Omega^{-1/2}M\Omega^{-1/2}u. \quad (C.2)$$

By (4.1) and the fact that  $Y = \tilde{X}\theta_0 + u$ ,  $\tilde{\theta}(\delta) = \theta_0 + \theta^*(\delta)$  where  $\theta^*(\delta) = [E(\tilde{X}'\Omega^{-1}\tilde{X})]^{-1}E(\tilde{X}'\Omega^{-1}u)$ . Then  $u(\tilde{\theta}(\delta)) = Y - \tilde{X}\tilde{\theta}(\delta) = u - \tilde{X}\theta^*(\delta)$ . By (4.2) and using the expression for  $\theta^*(\delta)$ , we have

$$\begin{aligned} \hat{\sigma}_v^2(\delta) &= \frac{1}{nT}E\{[u - \tilde{X}\theta^*(\delta)]'\Omega^{-1}[u - \tilde{X}\theta^*(\delta)]\} \\ &= \frac{1}{nT}E(u'\Omega^{-1}u) + \frac{1}{nT}\theta^*(\delta)'E(\tilde{X}'\Omega^{-1}\tilde{X})\theta^*(\delta) - \frac{2}{nT}\theta^*(\delta)'E(\tilde{X}'\Omega^{-1}u) \\ &= \frac{1}{nT}\sigma_{v0}^2\text{tr}(\Omega^{-1}\Omega_0) - \frac{1}{nT}[E(\tilde{X}'\Omega^{-1}u)]'[E(\tilde{X}'\Omega^{-1}\tilde{X})]^{-1}E(\tilde{X}'\Omega^{-1}u), \end{aligned} \quad (C.3)$$

where recall  $\Omega_0 \equiv \Omega(\delta_0)$  and  $\Omega(\delta)$  is defined in (3.2). Combining (C.2)-(C.3) yields

$$\begin{aligned} \hat{\sigma}_v^2(\delta) - \hat{\sigma}_v^2(\delta_0) &= \frac{1}{nT}[u'\Omega^{-1}u - \sigma_{v0}^2\text{tr}(\Omega^{-1}\Omega_0)] - \frac{1}{nT}u'\Omega^{-1/2}P\Omega^{-1/2}u \\ &\quad + \frac{1}{nT}[E(\tilde{X}'\Omega^{-1}u)]'[E(\tilde{X}'\Omega^{-1}\tilde{X})]^{-1}E(\tilde{X}'\Omega^{-1}u) \\ &= \frac{1}{nT}\text{tr}[\Omega^{-1}(uu' - \sigma_{v0}^2\Omega_0)] \\ &\quad - \{Q_{xu}(\delta)'Q_{xx}(\delta)^{-1}Q_{xu}(\delta) - E[Q_{xu}(\delta)']\{E[Q_{xx}(\delta)]\}^{-1}E[Q_{xu}(\delta)]\} \\ &\equiv \Pi_{n1}(\delta) - \Pi_{n2}(\delta), \text{ say,} \end{aligned}$$

where  $P = I_{nT} - M$ ,  $Q_{xx}(\delta) = \frac{1}{nT}\tilde{X}'\Omega^{-1}\tilde{X}$ , and  $Q_{xu}(\delta) = \frac{1}{nT}\tilde{X}'\Omega^{-1}u$ .

For  $\Pi_{n1}(\delta)$ , we can show that  $E[\Pi_{n1}(\delta)] = 0$  and  $E[\Pi_{n1}(\delta)]^2 = O(n^{-1})$  as in the proof of Lemma B.5. So the pointwise convergence of  $\Pi_{n1}(\delta)$  to 0 follows by Chebyshev inequality. The uniform convergence of  $\Pi_{n1}(\delta)$  to 0 holds if we can show that  $\Pi_{n1}(\delta)$  is stochastic equicontinuous. To achieve this, we first show that  $\inf_{\delta \in \Delta} \lambda_{\min}(\Omega(\delta))$  is bounded away from 0:

$$\begin{aligned} \inf_{\delta \in \Delta} \lambda_{\min}(\Omega(\delta)) &\geq \inf_{\delta \in \Delta} \lambda_{\min}\{\phi_\mu(J_T \otimes I_n) + I_T \otimes [B(\lambda)'B(\lambda)]^{-1}\} \\ &\geq \inf_{\lambda \in \Lambda} \lambda_{\min}(I_T \otimes [B(\lambda)'B(\lambda)]^{-1}) \geq \inf_{\lambda \in \Lambda} \lambda_{\min}([B(\lambda)'B(\lambda)]^{-1}) \\ &\geq \left\{ \sup_{\lambda \in \Lambda} \lambda_{\max}[B(\lambda)'B(\lambda)] \right\}^{-1} \geq \bar{c}_\lambda^{-1} > 0 \end{aligned} \quad (C.4)$$

by Facts 8.16.20 and B.14.20 in Bernstein (2005) and Assumption G2(v). Now, let  $\delta, \bar{\delta} \in \Delta$ . By Cauchy-Schwarz inequality,

$$\begin{aligned} |\Pi_{n1}(\delta) - \Pi_{n1}(\bar{\delta})| &= \left| \frac{1}{nT}\text{tr}\{\Omega(\delta)^{-1}[\Omega(\delta) - \Omega(\bar{\delta})]\Omega(\bar{\delta})^{-1}(uu' - \sigma_{v0}^2\Omega_0)\} \right| \\ &\leq \frac{1}{nT}[\text{tr}\{\Omega(\delta)^{-1}[\Omega(\delta) - \Omega(\bar{\delta})]\Omega(\bar{\delta})^{-2}[\Omega(\delta) - \Omega(\bar{\delta})]\Omega(\delta)^{-1}\}]^{1/2} \|uu' - \sigma_{v0}^2\Omega_0\| \\ &\leq [\lambda_{\min}(\Omega(\bar{\delta}))]^{-2} \frac{1}{\sqrt{nT}} \|\Omega(\delta) - \Omega(\bar{\delta})\| \frac{1}{\sqrt{nT}} \|uu' - \sigma_{v0}^2\Omega_0\|. \end{aligned}$$

Straightforward moment calculations and Chebyshev inequality lead to  $\frac{1}{\sqrt{nT}} \|uu' - \sigma_{v0}^2\Omega_0\| = O_p(1)$ . In addition,  $\frac{1}{\sqrt{nT}} \|\Omega(\delta) - \Omega(\bar{\delta})\| \rightarrow 0$  as  $\|\delta - \bar{\delta}\| \rightarrow 0$ . Thus,  $\{\Pi_{n1}(\delta)\}$  is stochastically equicontinuous by Theorem 21.10 in Davidson (1994).

For  $\Pi_{n2}(\delta)$ , we decompose it as follows

$$\begin{aligned} \Pi_{n2}(\delta) &= \{Q_{xu}(\delta) - E[Q_{xu}(\delta)]\}'Q_{xx}(\delta)^{-1}Q_{xu}(\delta) \\ &\quad + \{E[Q_{xu}(\delta)]\}'Q_{xx}(\delta)^{-1}\{E[Q_{xx}(\delta)] - Q_{xx}(\delta)\}\{E[Q_{xx}(\delta)]\}^{-1}Q_{xu}(\delta) \\ &\quad + \{E[Q_{xu}(\delta)]\}'\{E[Q_{xx}(\delta)]\}^{-1}\{Q_{xu}(\delta) - E[Q_{xu}(\delta)]\} \\ &\equiv \Pi_{n2,1}(\delta) + \Pi_{n2,2}(\delta) + \Pi_{n2,3}(\delta), \text{ say.} \end{aligned}$$

By Assumption G1(v),  $\sup |\phi_\mu| \leq c_\phi$  for some  $c_\phi < \infty$ . Noting that by G2(v)

$$\begin{aligned} \sup_{\delta \in \Delta} \lambda_{\max}(\Omega(\delta)) &\leq \sup_{\delta \in \Delta} \lambda_{\max}\{\phi_\mu(J_T \otimes I_n) + I_T \otimes [B(\lambda)'B(\lambda)]^{-1}\} \\ &\leq \sup_{\phi_\mu} \{\phi_\mu \lambda_{\max}(J_T \otimes I_n) + \lambda_{\max}\{[B(\lambda)'B(\lambda)]^{-1}\}\} \\ &\leq c_\phi T + \{\inf_{\lambda \in \Lambda} \lambda_{\min}[B(\lambda)'B(\lambda)]\}^{-1} \leq c_\phi T + \underline{\epsilon}_\lambda^{-1} < \infty, \end{aligned} \quad (\text{C.5})$$

we have  $\inf_{\delta \in \Delta} \lambda_{\min}(Q_{xx}(\delta)) \geq [\sup_{\delta \in \Delta} \lambda_{\max}(\Omega(\delta))]^{-1} \lambda_{\min}(\frac{1}{nT} \tilde{X}' \tilde{X}) \geq (c_\phi T + \underline{\epsilon}_\lambda^{-1})^{-1} \lambda_{\min}(\frac{1}{nT} \tilde{X}' \tilde{X})$ . This implies that  $\sup_{\delta \in \Delta} \|Q_{xx}(\delta)^{-1}\| = O_p(1)$  by Assumption R(iv). It is straightforward to show that  $Q_{xu}(\delta) - E[Q_{xu}(\delta)] = o_p(1)$  uniformly in  $\delta$  by Chebyshev inequality and the arguments for stochastic equicontinuity. In addition,  $E[Q_{xu}(\delta)] = O(1)$  and  $Q_{xu}(\delta) = O_p(1)$  uniformly in  $\delta$ . Consequently,

$$\begin{aligned} |\Pi_{n2,1}(\delta)| &\leq \|Q_{xu}(\delta) - E[Q_{xu}(\delta)]\| \|Q_{xx}(\delta)^{-1}\| \|Q_{xu}(\delta)\| \\ &= o_p(1) O_p(1) O_p(1) = o_p(1) \text{ uniformly in } \delta. \end{aligned}$$

By the same token, we can show that  $\Pi_{n2,s}(\delta) = o_p(1)$  uniformly in  $\delta$  for  $s = 2, 3$ . It follows that  $\Pi_{n2}(\delta) = o_p(1)$  uniformly in  $\delta$ . Hence  $\sup_{\delta \in \Delta} |\hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta)| = o_p(1)$  as desired.

To show (ii), we follow Lee (2002) and Yu et al. (2008) and first define an auxiliary process

$$Y^a = \rho Y_{-1}^a + X\beta + Z\gamma + U^a, \quad (\text{C.6})$$

where  $U^a \sim N(0, \sigma_v^2 \Omega)$  with  $\Omega = \Omega(\delta)$  and is independent of  $(X, Z)$ ,  $Y_{-1}^a$  and  $Y^a$  are analogously defined as  $Y_{-1}$  and  $Y$ , and the superscript  $a$  signifies that the process is an auxiliary one. Apparently, if  $u$  were normally distributed in (3.1), then one could simply set  $U^a$  as  $u$ , in which case  $(Y_{-1}^a, Y^a)$  would reduce to  $(Y_{-1}, Y)$ . As before, the true value of  $(\theta, \sigma_v^2, \delta)$  is given by  $(\theta_0, \sigma_{v0}^2, \delta_0)$ . The exact log-likelihood function of the above auxiliary process is given by

$$\log L_n^{r,a}(\theta, \sigma_v^2, \delta) = -\frac{nT}{2} \log(2\pi) - \frac{nT}{2} \log(\sigma_v^2) - \frac{1}{2} \log |\Omega| - \frac{1}{2\sigma_v^2} U^a(\theta)' \Omega^{-1} U^a(\theta) \quad (\text{C.7})$$

where  $U^a(\theta) = Y^a - \rho Y_{-1}^a - X\beta - Z\gamma$ . Let  $E^a$  denote expectation under this auxiliary process. By Jensen inequality,

$$0 = \log E^a \left( \frac{L_n^{r,a}(\theta, \sigma_v^2, \delta)}{L_n^{r,a}(\theta_0, \sigma_{v0}^2, \delta_0)} \right) \geq E^a \left[ \log \left( \frac{L_n^{r,a}(\theta, \sigma_v^2, \delta)}{L_n^{r,a}(\theta_0, \sigma_{v0}^2, \delta_0)} \right) \right].$$

That is,  $E^a [\log L_n^{r,a}(\theta, \sigma_v^2, \delta)] \leq E^a [\log L_n^{r,a}(\theta_0, \sigma_{v0}^2, \delta_0)]$ . Observe that  $\mathcal{L}_c^{r*}(\delta) = \max_{\theta, \sigma_v^2} E^a [\log L_n^{r,a}(\theta, \sigma_v^2, \delta)]$  and

$$\begin{aligned} E_a [\log L_n^{r,a}(\theta_0, \sigma_{v0}^2, \delta_0)] &= -\frac{nT}{2} \log(2\pi) - \frac{nT}{2} \log(\sigma_{v0}^2) - \frac{1}{2} \log |\Omega_0| - \frac{1}{2\sigma_{v0}^2} \text{tr}(\Omega_0^{-1} E[U^a(\theta_0) U^a(\theta_0)']) \\ &= -\frac{nT}{2} [\log(2\pi) + 1] - \frac{nT}{2} \log(\sigma_{v0}^2) - \frac{1}{2} \log |\Omega_0| = \mathcal{L}_c^{r*}(\delta_0) \end{aligned}$$

where we have used the fact that  $\tilde{\sigma}_v^2(\delta_0) = \frac{\sigma_{v0}^2}{nT} \text{tr}(\Omega_0^{-1} \Omega_0) = \sigma_{v0}^2$  by (C.3) and Lemma B.6 (see also the remark after (4.3)). It follows that

$$\mathcal{L}_c^{r*}(\delta) \leq E^a [\log L_n^{r,a}(\theta_0, \sigma_{v0}^2, \delta_0)] = \mathcal{L}_c^{r*}(\delta_0) \text{ for any } \delta \in \Delta. \quad (\text{C.8})$$

Next we show that  $\frac{1}{nT}\mathcal{L}_c^{r*}(\delta)$  is uniformly equicontinuous on  $\Delta$  by showing the uniform equicontinuity of  $\frac{1}{nT}\log|\Omega(\delta)|$  and  $\log[\tilde{\sigma}_v^2(\delta)]$  on  $\Delta$ . Let  $\delta_1$  and  $\delta_2$  be in  $\Delta$ . By the mean value theorem,  $\log|\Omega(\delta_1)| - \log|\Omega(\delta_2)| = (\frac{\partial}{\partial\bar{\delta}}\log|\Omega(\bar{\delta})|)(\delta_1 - \delta_2)$ , where  $\bar{\delta} = (\bar{\lambda}, \bar{\phi}_\mu)'$  lies elementwise between  $\delta_1$  and  $\delta_2$ . Note that

$$\frac{1}{nT}\frac{\partial}{\partial\bar{\lambda}}\log|\Omega(\bar{\delta})| = \frac{1}{nT}\text{tr}[\Omega(\bar{\delta})^{-1}(I_T \otimes A(\bar{\lambda}))]$$

where  $A(\bar{\lambda})$  is  $A = A(\lambda)$  evaluated at  $\bar{\lambda}$ . By (C.4) and the fact that  $\text{tr}(C_1 C_2) \leq \lambda_{\max}(C_1)\text{tr}(C_2)$  for any symmetric matrix  $C_1$  and positive semidefinite matrix  $C_2$ ,

$$\frac{1}{nT}|\text{tr}(\Omega^{-1}(I_T \otimes A))| \leq \frac{1}{nT}[\lambda_{\min}(\Omega)]^{-1}\text{tr}(I_T \otimes A) \leq \bar{c}_\lambda \frac{1}{n}\text{tr}(A) = O(1) \text{ uniformly on } \Delta.$$

It follows that  $\frac{1}{nT}\frac{\partial}{\partial\bar{\lambda}}\log|\Omega(\bar{\delta})| = O(1)$ . Similarly, and  $\frac{1}{nT}\frac{\partial}{\partial\phi_\mu}\log|\Omega(\bar{\delta})| = \text{tr}(\Omega(\bar{\delta})^{-1}(J_T \otimes I_n)) \leq \bar{c}_\lambda = O(1)$ . Thus  $\log|\Omega(\delta)|$  is uniformly equicontinuous in  $\delta$  on  $\Delta$ .

To show that  $\log[\tilde{\sigma}_v^2(\delta)]$  is uniformly equicontinuous on  $\Delta$ , it suffices to show that  $\tilde{\sigma}_v^2(\delta)$  is uniformly equicontinuous and uniformly bounded away from zero on  $\Delta$ . Observing that

$$\begin{aligned} \tilde{\sigma}_v^2(\delta) &= \frac{1}{nT}E[u(\tilde{\theta}(\delta))'\Omega^{-1}u(\tilde{\theta}(\delta))] \\ &= \frac{1}{nT}E(u'\Omega^{-1}u) + \frac{1}{nT}[\tilde{\theta}(\delta) - \theta_0]'E(\tilde{X}'\Omega^{-1}\tilde{X})[\tilde{\theta}(\delta) - \theta_0] + \frac{2}{nT}[\tilde{\theta}(\delta) - \theta_0]'E(\tilde{X}'\Omega^{-1}u) \\ &\equiv \tilde{\sigma}_{v1}^2(\delta) + \tilde{\sigma}_{v2}^2(\delta) + \tilde{\sigma}_{v3}^2(\delta), \text{ say,} \end{aligned}$$

the uniform equicontinuity of  $\tilde{\sigma}_v^2(\delta)$  follows from that of the three terms on the right hand side of the last equation. Note that

$$\begin{aligned} \left|\frac{\partial}{\partial\bar{\lambda}}\tilde{\sigma}_{v1}^2(\delta)\right| &= \frac{1}{nT}E[u'\Omega^{-1}(I_T \otimes A)\Omega^{-1}u] = \frac{\sigma_{v0}^2}{nT}\text{tr}[(I_T \otimes A)\Omega^{-1}\Omega_0\Omega^{-1}] \\ &\leq \sigma_{v0}^2\lambda_{\max}(\Omega^{-1}\Omega_0\Omega^{-1})\frac{1}{nT}\text{tr}(I_T \otimes A) \\ &\leq \sigma_{v0}^2[\lambda_{\min}(\Omega)]^{-2}\lambda_{\max}(\Omega_0)\frac{1}{n}\text{tr}(A) = O(1) \text{ uniformly in } \delta \end{aligned}$$

by (C.4), (C.5), and the fact that  $\frac{1}{n}\text{tr}(A) = O(1)$  uniformly in  $\lambda$  under Assumption G2. Similarly,

$$\begin{aligned} \left|\frac{\partial}{\partial\phi_\mu}\tilde{\sigma}_{v1}^2(\delta)\right| &= \frac{1}{nT}E[u'\Omega^{-1}(J_T \otimes I_n)\Omega^{-1}u] = \frac{\sigma_{v0}^2}{nT}\text{tr}[(J_T \otimes I_n)\Omega^{-1}\Omega_0\Omega^{-1}] \\ &\leq \sigma_{v0}^2\lambda_{\max}(\Omega^{-1}\Omega_0\Omega^{-1})\frac{1}{nT}\text{tr}(J_T \otimes I_n) \\ &= \sigma_{v0}^2[\lambda_{\min}(\Omega)]^{-2}\lambda_{\max}(\Omega_0) = O(1) \text{ uniformly in } \delta. \end{aligned}$$

Then by the mean value argument, we can show that  $\tilde{\sigma}_{v1}^2(\delta)$  is uniformly equicontinuous in  $\delta$  on  $\Delta$ . Analogously, we can show that  $\tilde{\theta}(\delta)$  and  $E(\tilde{X}'\Omega^{-1}\tilde{X})$  are uniformly equicontinuous on  $\Delta$ , which implies that  $\tilde{\sigma}_{v2}^2(\delta)$  and  $\tilde{\sigma}_{v3}^2(\delta)$  are uniformly equicontinuous on  $\Delta$ . Thus we can conclude that  $\tilde{\sigma}_v^2(\delta)$  is uniformly equicontinuous on  $\Delta$ . To show that  $\tilde{\sigma}_v^2(\delta)$  is uniformly bounded away from zero, we make its dependence on  $n$  explicit and write it as  $\tilde{\sigma}_{v,n}^2(\delta)$ . We establish the claim by a counter argument. Suppose that  $\tilde{\sigma}_{v,n}^2(\delta)$  is not uniformly bounded away from zero on  $\Delta$ . Then there exists a sequence  $\{\delta_n\}$  in  $\Delta$  such that  $\lim_{n \rightarrow \infty} \tilde{\sigma}_{v,n}^2(\delta_n) = 0$ . By (C.8), we have  $\frac{1}{nT}[\mathcal{L}_c^{r*}(\delta) - \mathcal{L}_c^{r*}(\delta_0)] \leq 0$  for all  $\delta$ , i.e.,

$$-\log[\tilde{\sigma}_v^2(\delta)] \leq -\log[\tilde{\sigma}_v^2(\delta_0)] + \frac{1}{nT}\{\log|\Omega(\delta)| - \log|\Omega_0|\}.$$

By (C.4) and (C.5) and the mean value theorem, we can readily show that  $\frac{1}{nT}\{\log|\Omega(\delta)| - \log|\Omega_0|\} = O(1)$  uniformly on  $\Delta$ . This implies that  $-\log[\tilde{\sigma}_v^2(\delta)]$  is bounded above, a contradiction. Therefore we can conclude that  $\tilde{\sigma}_{v,n}^2(\delta)$  is uniformly bounded away from zero on  $\Delta$ .



Now, the identification uniqueness follows by contradiction. Using  $\tilde{\sigma}_v^2(\delta_0) = \sigma_{v0}^2$  again we have

$$\begin{aligned} \frac{1}{nT}[\mathcal{L}_c^{r*}(\delta) - \mathcal{L}_c^{r*}(\delta_0)] &= \frac{1}{2nT} \{ \log |\Omega_0| - \log |\Omega(\delta)| \} + \frac{1}{2} \{ \log [\tilde{\sigma}_v^2(\delta_0)] - \log [\tilde{\sigma}_v^2(\delta)] \} \\ &= \frac{1}{2nT} \{ \log |\sigma_{v0}^2 \Omega_0| - \log |\tilde{\sigma}_v^2(\delta) \Omega(\delta)| \}. \end{aligned} \quad (\text{C.9})$$

Suppose that the identification uniqueness condition does not hold. Then there exists an  $\epsilon > 0$  and a sequence  $\{\delta_n\}$  in  $N_\epsilon^c(\delta_0)$  such that

$$\lim_{n \rightarrow \infty} \frac{1}{nT} [\mathcal{L}_{c,n}^{r*}(\delta_n) - \mathcal{L}_{c,n}^{r*}(\delta_0)] = 0$$

where we write  $\mathcal{L}_{c,n}^{r*}(\cdot)$  for  $\mathcal{L}_c^{r*}(\cdot)$  to make its dependence on  $n$  explicit. By the compactness of  $N_\epsilon^c(\delta_0)$ , there exists a convergent subsequence  $\{\delta_{n_k}\}$  of  $\{\delta_n\}$  with the limit  $\delta_+$  of  $\delta_{n_k}$  being in  $N_\epsilon^c(\delta_0)$ . This implies that  $\delta_+ \neq \delta_0$ . Furthermore,  $\lim_{n \rightarrow \infty} \frac{1}{n_k T} [\mathcal{L}_{c,n_k}^{r*}(\delta_+) - \mathcal{L}_{c,n_k}^{r*}(\delta_0)] = 0$  by the uniform equicontinuity of  $\frac{1}{nT} \mathcal{L}_{c,n}^{r*}(\delta)$ . But this contradicts to Assumption R(iv) as it is equivalent to that  $\lim_{n \rightarrow \infty} \frac{1}{nT} [\mathcal{L}_{c,n}^{r*}(\delta) - \mathcal{L}_{c,n}^{r*}(\delta_0)] \neq 0$  for any  $\delta \neq \delta_0$ . This completes the proof of the theorem. ■

**Proof of Theorem 4.2.** By Taylor series expansion,

$$0 = \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}^r(\hat{\psi})}{\partial \psi} = \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}^r(\psi_0)}{\partial \psi} + \frac{1}{nT} \frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \psi \partial \psi'} \sqrt{nT}(\hat{\psi} - \psi_0),$$

where elements of  $\bar{\psi} = (\bar{\theta}', \bar{\sigma}_v^2, \bar{\delta})'$  lie in the segment joining the corresponding elements of  $\hat{\psi}$  and  $\psi_0$  and  $\bar{\delta} = (\bar{\lambda}, \bar{\phi}_\mu)'$ . Thus

$$\sqrt{nT}(\hat{\psi} - \psi_0) = - \left[ \frac{1}{nT} \frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \psi \partial \psi'} \right]^{-1} \frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}^r(\psi_0)}{\partial \psi}.$$

By Theorem 4.1,  $\hat{\psi} \xrightarrow{p} \psi_0$ , and thus  $\bar{\psi} \xrightarrow{p} \psi_0$ . It suffices to show that: (i)  $\frac{1}{nT} \frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \psi \partial \psi'} - \frac{1}{nT} \frac{\partial^2 \mathcal{L}^r(\psi_0)}{\partial \psi \partial \psi'} = o_p(1)$ , (ii)  $\frac{1}{nT} \frac{\partial^2 \mathcal{L}^r(\psi_0)}{\partial \psi \partial \psi'} \xrightarrow{p} H_r$ , and (iii)  $\frac{1}{\sqrt{nT}} \frac{\partial \mathcal{L}^r(\psi_0)}{\partial \psi} \xrightarrow{d} N(0, \Gamma_r)$ . (ii) and (iii) follow from Lemmas B.7 and B.8, respectively. We are left to show (i).

With the expression of  $\frac{\partial^2 \mathcal{L}^r(\psi)}{\partial \psi \partial \psi'}$  given in Appendix A, it suffices to show that  $\frac{1}{nT} \frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \omega \partial \omega'} - \frac{1}{nT} \frac{\partial^2 \mathcal{L}^r(\psi_0)}{\partial \omega \partial \omega'} = o_p(1)$  for  $\omega, \varpi = \theta, \sigma_v^2, \lambda$ , and  $\phi_\mu$ . We do this only for the cases of  $(\omega, \varpi) = (\theta, \theta)$ ,  $(\theta, \sigma_v^2)$ , and  $(\sigma_v^2, \sigma_v^2)$  as the other cases can be shown analogously. First, write

$$-\frac{1}{nT} \left[ \frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \theta \partial \theta'} - \frac{\partial^2 \mathcal{L}^r(\psi_0)}{\partial \theta \partial \theta'} \right] = \left( \frac{1}{\bar{\sigma}_v^2} - \frac{1}{\sigma_{v0}^2} \right) \frac{\tilde{X}' \Omega(\bar{\delta})^{-1} \tilde{X}}{nT} + \frac{1}{nT \sigma_{v0}^2} \tilde{X}' [\Omega(\bar{\delta})^{-1} - \Omega_0^{-1}] \tilde{X}. \quad (\text{C.10})$$

Noting that  $\bar{\sigma}_v^2 - \sigma_{v0}^2 = o_p(1)$  by Theorem 4.1 and  $(nT)^{-1} \tilde{X}' \Omega(\bar{\delta})^{-1} \tilde{X} = O_p(1)$ , the first term on the right hand side of the last expression is  $o_p(1)$ . For the second term, we first show that

$$\lambda_{\max}[\Omega_0 - \Omega(\bar{\delta})] = O_p(\|\bar{\delta} - \delta_0\|). \quad (\text{C.11})$$

To see this, write  $\Omega_0 - \Omega(\bar{\delta}) = (\phi_{\mu 0} - \bar{\phi}_\mu)(J_T \otimes I_n) + r_n(\bar{\lambda})$ , where  $r_n(\lambda) = I_T \otimes \{[B(\lambda_0)' B(\lambda_0)]^{-1} - [B(\lambda)' B(\lambda)]^{-1}\}$  is a symmetric matrix. By the repeated use of the fact that

$$\lambda_{\max}(A \otimes C) \leq \lambda_{\max}(A) \lambda_{\max}(C) \quad (\text{C.12})$$

for any two real symmetric matrices [see, e.g., Fact 8.16.20 of Bernstein (2005)], we have

$$\begin{aligned} \lambda_{\max}[r_n(\bar{\lambda})] &\leq \lambda_{\max}\{[B(\lambda_0)' B(\lambda_0)]^{-1} - [B(\bar{\lambda})' B(\bar{\lambda})]^{-1}\} \\ &= \lambda_{\max}([B(\lambda_0)' B(\lambda_0)]^{-1} [B(\bar{\lambda})' B(\bar{\lambda}) - B(\lambda_0)' B(\lambda_0)] [B(\bar{\lambda})' B(\bar{\lambda})]^{-1}) \\ &\leq \left\{ \inf_{\lambda \in \Lambda} \lambda_{\min}[B(\lambda)' B(\lambda)] \right\}^{-2} \lambda_{\max}[B(\bar{\lambda})' B(\bar{\lambda}) - B(\lambda_0)' B(\lambda_0)] = O_p(\bar{\lambda} - \lambda_0) \end{aligned}$$

where the last equality follows from Assumption G2 and the fact that

$$\begin{aligned}\lambda_{\max}[B(\bar{\lambda})'B(\bar{\lambda}) - B(\lambda_0)'B(\lambda_0)] &= \lambda_{\max}[(\lambda_0 - \bar{\lambda})(W' + W) + (\bar{\lambda}^2 - \lambda_0^2)W'W] \\ &\leq |\bar{\lambda} - \lambda_0|\lambda_{\max}(W' + W) + (\bar{\lambda}^2 - \lambda_0^2)\lambda_{\max}(W'W) \\ &= O_p(\bar{\lambda} - \lambda_0)\end{aligned}$$

under Assumption G2. Noting that  $\lambda_{\max}(J_T \otimes I_n) = T$ , we can apply the fact that

$$\lambda_{\max}(A + C) \leq \lambda_{\max}(A) + \lambda_{\max}(C) \quad (\text{C.13})$$

to obtain  $\lambda_{\max}[\Omega_0 - \Omega(\bar{\delta})] \leq T|\phi_{\mu 0} - \bar{\phi}_{\mu}| + \lambda_{\max}(r_n(\bar{\lambda})) = O_p(\|\bar{\delta} - \delta_0\|)$ . Thus (C.11) follows. Let  $c$  be an arbitrary column vector in  $\mathbb{R}^{p+q+1}$ . Then by Cauchy-Schwarz inequality, (C.4), and (C.11)

$$\begin{aligned}& \frac{1}{n}|c'\tilde{X}'[\Omega(\bar{\delta})^{-1} - \Omega_0^{-1}]\tilde{X}c| \\ &= \frac{1}{n}|c'\tilde{X}'\Omega(\bar{\delta})^{-1}[\Omega_0 - \Omega(\bar{\delta})]\Omega_0^{-1}\tilde{X}c| \\ &\leq \frac{1}{n}\{c'\tilde{X}'\Omega(\bar{\delta})^{-1}[\Omega_0 - \Omega(\bar{\delta})][\Omega_0 - \Omega(\bar{\delta})]\Omega(\bar{\delta})^{-1}\tilde{X}c\}^{1/2}[c'\tilde{X}'\Omega_0^{-1}\Omega_0^{-1}\tilde{X}c]^{1/2} \\ &\leq \lambda_{\max}[\Omega_0 - \Omega(\bar{\delta})][\lambda_{\min}(\Omega(\bar{\delta}))]^{-1}[\lambda_{\min}(\Omega_0)]^{-1}\frac{1}{n}\|\tilde{X}c\|^2 = O_p(\|\bar{\delta} - \delta_0\|) = o_p(1).\end{aligned} \quad (\text{C.14})$$

It follows that the second term on the right hand side of (C.10) is  $o_p(1)$ . Consequently,  $\frac{1}{nT}\frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \theta \partial \theta'} - \frac{1}{nT}\frac{\partial^2 \mathcal{L}^r(\psi_0)}{\partial \theta \partial \theta'} = o_p(1)$ .

Next we consider  $-\frac{1}{nT}\frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \theta \partial \sigma_v^2} + \frac{1}{nT}\frac{\partial^2 \mathcal{L}^r(\psi_0)}{\partial \theta \partial \sigma_v^2}$ . This term is equal to

$$\begin{aligned}& \frac{1}{nT\tilde{\sigma}_v^4}\tilde{X}'\Omega(\bar{\delta})^{-1}u(\bar{\theta}) - \frac{1}{nT\sigma_{v0}^4}\tilde{X}'\Omega_0^{-1}u \\ &= \left(\frac{1}{\tilde{\sigma}_v^4} - \frac{1}{\sigma_{v0}^4}\right)\frac{\tilde{X}'\Omega(\bar{\delta})^{-1}u(\bar{\theta})}{nT} + \frac{1}{\sigma_{v0}^4}\frac{\tilde{X}'[\Omega(\bar{\delta})^{-1} - \Omega_0^{-1}]u(\bar{\theta})}{nT} + \frac{1}{\sigma_{v0}^4}\frac{\tilde{X}'\Omega_0^{-1}[u(\bar{\theta}) - u]}{nT}.\end{aligned}$$

Using  $u(\bar{\theta}) = Y - \tilde{X}\bar{\theta} = u + \tilde{X}(\theta_0 - \bar{\theta})$ , we can readily show that  $\frac{1}{nT}\tilde{X}'\Omega(\bar{\delta})^{-1}u(\bar{\theta}) = O_p(1)$ , which implies that the first term in the last expression is  $o_p(1)$  by Theorem 4.1. The second term is  $o_p(1)$  by arguments analogous to those used above. The third term is  $\sigma_{v0}^{-4}(nT)^{-1}\tilde{X}'\Omega(\bar{\delta})^{-1}\tilde{X}(\theta_0 - \bar{\theta}) = O_p(1)\|\theta_0 - \bar{\theta}\| = o_p(1)$ . It follows that  $\frac{1}{nT}\frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \theta \partial \sigma_v^2} - \frac{1}{nT}\frac{\partial^2 \mathcal{L}^r(\psi_0)}{\partial \theta \partial \sigma_v^2} = o_p(1)$ . Now, write

$$-\frac{1}{nT}\left[\frac{\partial^2 \mathcal{L}^r(\bar{\psi})}{\partial \sigma_v^2 \partial \sigma_v^2} - \frac{\partial^2 \mathcal{L}^r(\psi_0)}{\partial \sigma_v^2 \partial \sigma_v^2}\right] = \left(\frac{1}{\sigma_v^6}u(\bar{\theta})'\Omega(\bar{\delta})^{-1}u(\bar{\theta}) - \frac{1}{\sigma_v^2}u'\Omega_0^{-1}u\right) + \frac{1}{2}\left(\frac{1}{\sigma_{v0}^4} - \frac{1}{\sigma_v^4}\right).$$

Clearly, the second term is  $o_p(1)$  by Theorem 4.1. We can use the decomposition  $u(\bar{\theta}) = u + \tilde{X}(\theta_0 - \bar{\theta})$  and the consistency of  $\bar{\psi}$  to show the first term is also  $o_p(1)$ . This completes the proof. ■

**Proof of Theorem 4.3.** As in the proof of Theorem 4.1, we prove the theorem by showing that

(i)  $\frac{1}{nT_1}[\mathcal{L}_c^{rr*}(\delta) - \mathcal{L}_c^{rr}(\delta)] \xrightarrow{p} 0$  uniformly in  $\delta \in \Delta$ , and (ii)  $\limsup_{n \rightarrow \infty} \max_{\delta \in N_\epsilon^c(\delta_0)} \frac{1}{nT_1}[\mathcal{L}_c^{rr*}(\delta) - \mathcal{L}_c^{rr*}(\delta_0)] < 0$  for any  $\epsilon > 0$ , where  $T_1 = T + 1$ .

By (3.14) and (4.6),  $\frac{2}{nT_1}[\mathcal{L}_c^{rr*}(\delta) - \mathcal{L}_c^{rr}(\delta)] = \ln \hat{\sigma}_v^2(\delta) - \ln \tilde{\sigma}_v^2(\delta)$ . To show (i), it suffices to show

$$\hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta) = o_p(1) \text{ uniformly on } \Delta \quad (\text{C.15})$$

provided that  $\tilde{\sigma}_v^2(\delta)$  is uniformly bounded away from zero. By the definition of  $\tilde{u}^*(\delta)$  below (3.13), we have  $\tilde{u}^*(\delta) = Y^*(\rho) - X^*(X^{*'}\Omega^{*-1}X^*)^{-1}X^{*'}\Omega^{*-1}Y^*(\rho) = \Omega^{*1/2}M^*\Omega^{*-1/2}Y^*(\rho)$  where  $M^* = I_{nT_1} -$

$\Omega^{*-1/2} X^* (X^{*'} \Omega^{*-1} X^*)^{-1} X^{*'} \Omega^{*-1/2}$  is a projection matrix. Observe that  $Y^*(\rho) = Y^*(\rho_0) + [Y^*(\rho) - Y^*(\rho_0)] = X^* \theta_0 + u^* + (\rho_0 - \rho) Y_{-1}^*$  where  $Y_{-1}^* = (0_{1 \times n}, Y_{-1}^*)'$ . This, in conjunction with the fact that  $M^* \Omega^{*-1/2} X^* = 0$ , implies that

$$\begin{aligned} \hat{\sigma}_v^2(\delta) &= \frac{1}{nT_1} \tilde{u}^*(\delta)' \Omega^{*-1} \tilde{u}^*(\delta) = \frac{1}{nT_1} Y^*(\rho)' \Omega^{*-1/2} M^* \Omega^{*-1/2} Y^*(\rho) \\ &= \frac{1}{nT_1} [u^* + (\rho_0 - \rho) Y_{-1}^*]' \Omega^{*-1/2} M^* \Omega^{*-1/2} [u^* + (\rho_0 - \rho) Y_{-1}^*]. \end{aligned} \quad (C.16)$$

By (4.4) and the above expression for  $Y^*(\rho)$ , we have

$$\tilde{\theta}(\delta) = [E(X^{*'} \Omega^{*-1} X^*)]^{-1} E[X^{*'} \Omega^{*-1} Y^*(\rho)] = \theta_0 - \theta^*(\delta),$$

where  $\theta^*(\delta) = (\rho - \rho_0) [E(X^{*'} \Omega^{*-1} X^*)]^{-1} E(X^{*'} \Omega^{*-1} Y_{-1}^*)$ . Then by the definition of  $u^*(\theta, \rho)$  after (3.12),

$$u^*(\tilde{\theta}(\delta), \rho) = Y^*(\rho) - X^* \tilde{\theta}(\delta) = X^* \theta^*(\delta) + u^* + (\rho_0 - \rho) Y_{-1}^*.$$

By (4.5),

$$\begin{aligned} \tilde{\sigma}_v^2(\delta) &= \frac{1}{nT_1} E\{[X^* \theta^*(\delta) + u^* + (\rho_0 - \rho) Y_{-1}^*]' \Omega^{*-1} [X^* \theta^*(\delta) + u^* + (\rho_0 - \rho) Y_{-1}^*]\} \\ &= \frac{1}{nT_1} E[v^*(\delta)] + \frac{1}{nT_1} \theta^*(\delta)' E(X^{*'} \Omega^{*-1} X^*) \theta^*(\delta) + \frac{2(\rho_0 - \rho)}{nT_1} \theta^*(\delta)' E(X^{*'} \Omega^{*-1} Y_{-1}^*) \\ &= \frac{1}{nT_1} E[v^*(\delta)] + \frac{(\rho_0 - \rho)}{nT_1} \theta^*(\delta)' E(X^{*'} \Omega^{*-1} Y_{-1}^*), \end{aligned} \quad (C.17)$$

where  $v^*(\delta) = [u^* + (\rho_0 - \rho) Y_{-1}^*]' \Omega^{*-1} [u^* + (\rho_0 - \rho) Y_{-1}^*]$ . Using (C.16)-(C.17), and  $\Omega^{*-1/2} M^* \Omega^{*-1/2} = \Omega^{*-1} - \Omega^{*-1} X^* (X^{*'} \Omega^{*-1} X^*)^{-1} X^{*'} \Omega^{*-1}$ , we have

$$\begin{aligned} \hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta) &= \frac{1}{nT_1} \{v^*(\delta) - E[v^*(\delta)]\} + Q_{xu}^*(\delta)' Q_{xx}^*(\delta)^{-1} Q_{xu}^*(\delta)' + 2(\rho_0 - \rho) Q_{xu}^*(\delta)' Q_{xx}^*(\delta)^{-1} Q_{xy-1}^*(\delta) \\ &\quad + (\rho_0 - \rho)^2 \{Q_{xy-1}^*(\delta)' Q_{xx}^*(\delta)^{-1} Q_{xy-1}^*(\delta) - E[Q_{xy-1}^*(\delta)'] \{E[Q_{xx}^*(\delta)]\}^{-1} E[Q_{xy-1}^*(\delta)]\} \\ &\equiv \Pi_{n1}^*(\delta) + \Pi_{n2}^*(\delta) + 2(\rho_0 - \rho) \Pi_{n3}^*(\delta) + (\rho_0 - \rho)^2 \Pi_{n4}^*(\delta), \text{ say,} \end{aligned}$$

where  $Q_{xx}^*(\delta) = \frac{1}{nT_1} X^{*'} \Omega^{*-1} X^*$ ,  $Q_{xu}^*(\delta) = \frac{1}{nT_1} X^{*'} \Omega^{*-1} u^*$ , and  $Q_{xy-1}^*(\delta) = \frac{1}{nT_1} X^{*'} \Omega^{*-1} Y_{-1}^*$ . We prove (C.15) by showing that  $\Pi_{ns}^*(\delta) = o_p(1)$  uniformly in  $\delta$  for  $s = 1, 2, 3$ , and 4.

We can decompose  $\Pi_{n1}^*(\delta)$  as follows

$$\begin{aligned} \Pi_{n1}^*(\delta) &= \frac{1}{nT_1} [u^{*'} \Omega^{*-1} u^* - E(u^{*'} \Omega^{*-1} u^*)] + \frac{(\rho_0 - \rho)^2}{nT_1} [Y_{-1}^{*'} \Omega^{*-1} Y_{-1}^* - E(Y_{-1}^{*'} \Omega^{*-1} Y_{-1}^*)] \\ &\quad + \frac{2(\rho_0 - \rho)}{nT_1} [u^{*'} \Omega^{*-1} Y_{-1}^* - E(u^{*'} \Omega^{*-1} Y_{-1}^*)] \\ &\equiv \Pi_{n1,1}^*(\delta) + \Pi_{n1,2}^*(\delta) + \Pi_{n1,3}^*(\delta), \text{ say.} \end{aligned}$$

For  $\Pi_{n1,1}^*(\delta)$ , we can show that  $E[\Pi_{n1,1}^*(\delta)] = 0$  and  $E[\Pi_{n1,1}^*(\delta)]^2 = O(n^{-1})$  as in the proof of Lemma B.5. So the pointwise convergence of  $\Pi_{n1,1}^*(\delta)$  to 0 follows by Chebyshev inequality. The uniform convergence holds if we can show that  $\Pi_{n1,1}^*(\delta)$  is stochastic equicontinuous. Let  $\delta, \bar{\delta} \in \Delta$ . By Cauchy-Schwarz inequality,

$$\begin{aligned} &|\Pi_{n1,1}^*(\delta) - \Pi_{n1,1}^*(\bar{\delta})| \\ &= \left| \frac{1}{nT_1} \text{tr} \{ \Omega^*(\delta)^{-1} [\Omega^*(\bar{\delta}) - \Omega^*(\delta)] \Omega^*(\bar{\delta})^{-1} [u^* u^{*'} - E(u^* u^{*'})] \} \right| \\ &\leq \frac{1}{nT_1} \{ \text{tr} [\Omega^*(\delta)^{-1} (\Omega^*(\bar{\delta}) - \Omega^*(\delta)) \Omega^*(\bar{\delta})^{-2} (\Omega^*(\bar{\delta}) - \Omega^*(\delta)) \Omega^*(\delta)^{-1}] \}^{1/2} \|u^* u^{*'} - E(u^* u^{*'})\| \\ &\leq [\lambda_{\min}(\Omega^*(\bar{\delta}))]^{-2} \frac{1}{\sqrt{nT_1}} \|\Omega^*(\bar{\delta}) - \Omega^*(\delta)\| \frac{1}{\sqrt{nT_1}} \|u^* u^{*'} - E(u^* u^{*'})\|. \end{aligned}$$

Straightforward moment calculations and Chebyshev inequality lead to  $\frac{1}{\sqrt{nT_1}} \|u^* u^{*'} - E(u^* u^{*'})\| = O_p(1)$ . In addition,  $\frac{1}{\sqrt{nT_1}} \|\Omega^*(\bar{\delta}) - \Omega^*(\delta)\| \rightarrow 0$  as  $\|\delta - \bar{\delta}\| \rightarrow 0$ . Thus,  $\{\Pi_{n1,1}^*(\delta)\}$  is stochastically equicontinuous by Theorem 21.10 in Davidson (1994). Consequently,  $\Pi_{n1,1}^*(\delta) = o_p(1)$  uniformly in  $\delta$ . Similarly,  $\Pi_{n1,s}^*(\delta) = o_p(1)$  uniformly in  $\delta$  for  $s = 2, 3$ . It follows that  $\Pi_{n1}^*(\delta) = o_p(1)$  uniformly in  $\delta$ .

To show  $\Pi_{n2}^*(\delta) = o_p(1)$  uniformly in  $\delta$ , we first argue that  $\Omega^*(\delta)$  is positive definite uniformly in  $\delta$ , i.e.,  $\inf_{\delta \in \Delta} \lambda_{\min}(\Omega^*(\delta)) \geq \underline{c}^*$  for some  $\underline{c}^* > 0$ . Let  $\bar{u}^* = (a_m \mu', u')'$ . We have,

$$\bar{\Omega}^*(\delta) = E(\bar{u}^* \bar{u}^{*'}) = \begin{pmatrix} \phi_\mu a_m^2 I_n & \phi_\mu a_m (\iota_T' \otimes I_n) \\ \phi_\mu a_m (\iota_T \otimes I_n) & \Omega \end{pmatrix},$$

which is positive semidefinite uniformly in  $\delta$ . By Theorem 8.4.11 in Bernstein (2005) and (C.4),  $\lambda_{\min}(\phi_\zeta I_n + b_m(B'B)^{-1}) \geq \phi_\zeta + b_m \lambda_{\min}((B'B)^{-1}) \geq \phi_\zeta + b_m \bar{c}_\lambda^{-2} > 0$  uniformly in  $\delta$  as  $\phi_\zeta$  is positive and bounded away from 0 and  $b_m > 0$ , implying that  $\phi_\zeta I_n + b_m(B'B)^{-1}$  is positive definite uniformly in  $\delta$ . Noting  $\Omega^*(\delta)$  is equal to  $\bar{\Omega}^*(\delta)$  with its upper-left  $(n, n)$ -submatrix added by a uniformly positive definite matrix  $\phi_\zeta I_n + b_m(B'B)^{-1}$ , we can apply Fact 8.9.19 in Bernstein (2005) to conclude that  $\Omega^*(\delta)$  is positive definite uniformly in  $\delta$ . Similarly, we can readily show that

$$\begin{aligned} \sup_{\delta \in \Delta} \lambda_{\max}(\Omega^*(\delta)) &\leq \sup_{\delta \in \Delta} \lambda_{\max}(\bar{\Omega}^*(\delta)) + \sup_{\delta \in \Delta} \lambda_{\max}(\phi_\zeta I_n + b_m(B'B)^{-1}) \\ &\leq \sup_{\delta \in \Delta} \lambda_{\max}(\bar{\Omega}^*(\delta)) + \sup_{\delta \in \Delta} \phi_\zeta + b_m(\lambda_{\min}(B'B))^{-1} \leq \bar{c}^*, \text{ say.} \end{aligned}$$

Next, write

$$\begin{aligned} \frac{1}{nT_1} X^{*'} X^* &= \frac{1}{nT_1} \begin{pmatrix} X'X & X'Z & 0_{p \times k} \\ Z'X & Z'Z & 0_{q \times k} \\ \tilde{\mathbf{x}}'x_0 & \tilde{\mathbf{x}}'z_m(\rho) & \tilde{\mathbf{x}}'\tilde{\mathbf{x}} \end{pmatrix} + \frac{1}{nT_1} \begin{pmatrix} x_0'x_0 & x_0'z_m(\rho) & x_0'\tilde{\mathbf{x}} \\ z_m(\rho)'x_0 & z_m(\rho)'z_m(\rho) & z_m(\rho)'\tilde{\mathbf{x}} \\ 0_{k \times p} & 0_{k \times q} & 0_{k \times k} \end{pmatrix} \\ &\equiv A_1(\rho) + A_2(\rho), \text{ say.} \end{aligned}$$

Noting that  $A_1(\rho)$  is a block triangular matrix, its eigenvalues are those of the square matrices on the diagonal direction. By Assumption R\*(iv), the minimum of these eigenvalues are bounded away from 0, say by  $\underline{c}_{xx}$ , uniformly in  $\rho$ . Similarly, the minimum eigenvalues of  $A_2(\rho)$  is 0 uniformly in  $\rho$ . It follows that  $\inf_\rho \lambda_{\min}(\frac{1}{nT_1} X^{*'} X^*) \geq \inf_\rho [\lambda_{\min}(A_1(\rho)) + \lambda_{\min}(A_2(\rho))] \geq \underline{c}_{xx} > 0$ . Consequently,

$$\inf_{\delta \in \Delta} \lambda_{\min}(Q_{xx}^*(\delta)) = \inf_{\delta \in \Delta} \lambda_{\min}\left(\frac{1}{nT_1} X^{*'} \Omega^{*-1} X^*\right) \geq \bar{c}^{*-1} \inf_\rho \lambda_{\min}\left(\frac{1}{nT_1} X^{*'} X^*\right) \geq \bar{c}^{*-1} \underline{c}_{xx} > 0. \quad (\text{C.18})$$

Next, noting that  $E[Q_{xu}^*(\delta)] = 0$  and  $\text{Var}(Q_{xu}^*(\delta)) = O(n^{-1})$ , we have  $Q_{xu}^*(\delta) = o_p(1)$  by Chebyshev inequality. In addition, it is straightforward to show that  $Q_{xu}^*(\delta)$  is stochastic equicontinuous. So  $Q_{xu}^*(\delta) = o_p(1)$  uniformly in  $\delta$ . We have,  $|\Pi_{n2}^*(\delta)| \leq [\inf_{\delta \in \Delta} \lambda_{\min}(Q_{xx}^*(\delta))]^{-1} \|Q_{xu}^*(\delta)\|^2 = o_1(1)$  uniformly in  $\delta$ .

For  $\Pi_{n3}^*(\delta)$ , we have  $\Pi_{n3}^*(\delta) \leq \|Q_{xu}^*(\delta)\| \|Q_{xx}^*(\delta)^{-1}\| \|Q_{xy-1}^*(\delta)\| = o_p(1)$  uniformly in  $\delta$  as one can readily show that  $Q_{xy-1}^*(\delta) = o_p(1)$  uniformly in  $\delta$ .

For  $\Pi_{n4}^*(\delta)$ , we have

$$\begin{aligned} \Pi_{n4}^*(\delta) &= \{Q_{xy-1}^*(\delta) - E[Q_{xy-1}^*(\delta)]\}' Q_{xx}^*(\delta)^{-1} Q_{xy-1}^*(\delta) \\ &\quad + E[Q_{xy-1}^*(\delta)]' Q_{xx}^*(\delta)^{-1} \{E[Q_{xx}^*(\delta)] - Q_{xx}^*(\delta)\} \{E[Q_{xx}^*(\delta)]\}^{-1} Q_{xy-1}^*(\delta) \\ &\quad + E[Q_{xy-1}^*(\delta)]' E[Q_{xx}^*(\delta)] \{Q_{xy-1}^*(\delta) - E[Q_{xy-1}^*(\delta)]\} \\ &\equiv \Pi_{n4,1}^*(\delta) + \Pi_{n4,2}^*(\delta) + \Pi_{n4,3}^*(\delta), \text{ say.} \end{aligned}$$

We can readily show that  $Q_{xy-1}^*(\delta) - E[Q_{xy-1}^*(\delta)] = o_p(1)$  uniformly in  $\delta$  by Chebyshev inequality and the arguments of stochastic equicontinuity. This, in conjunction with (C.18) and the fact that  $Q_{xy-1}^*(\delta) = O_p(1)$  uniformly in  $\delta$ , implies that  $\Pi_{n4,1}^*(\delta) = o_p(1)$  uniformly in  $\delta$ . Similarly, we can show that  $\Pi_{n4,s}^*(\delta) = o_p(1)$  uniformly in  $\delta$  for  $s = 2, 3$ . Thus  $\Pi_{n4}^*(\delta) = o_p(1)$  uniformly in  $\delta$ . This completes the proof of (i).

The proof of (ii) is analogous to that of part (ii) in the proof of Theorem 4.1 and we only sketch the major differences. First, by the use of an auxiliary process  $Y^a$  that has the error term  $U^a$  being  $N(0, \sigma_v^2 \Omega^*(\delta))$  and independent of  $(X, Z)$ , we can apply Jensen inequality and the fact that  $\tilde{\sigma}_v^2(\delta_0) = \frac{1}{nT_1} E[u^{*'} \Omega_0^{*-1} u^*] = \frac{\sigma_{v0}^2}{nT} \text{tr}(\Omega_0^{*-1} \Omega_0^*) = \sigma_{v0}^2$  by (C.17) to show that

$$\mathcal{L}_c^{rr*}(\delta) \leq \mathcal{L}_c^{rr*}(\delta_0) \text{ for any } \delta \in \Delta. \quad (\text{C.19})$$

As before, we show that  $\frac{1}{nT_1} \mathcal{L}_c^{rr*}(\delta)$  is uniformly equicontinuous on  $\Delta$  by showing the uniform equicontinuity of  $\frac{1}{nT} \log |\Omega^*(\delta)|$  and  $\log [\tilde{\sigma}_v^2(\delta)]$  on  $\Delta$ . Let  $\delta_1$  and  $\delta_2$  be in  $\Delta$ . By the mean value theorem,  $\log |\Omega^*(\delta_1)| - \log |\Omega^*(\delta_2)| = (\partial \log |\Omega^*(\bar{\delta})| / \partial \delta)' (\delta_1 - \delta_2)$ , where  $\bar{\delta}$  lies elementwise between  $\delta_1$  and  $\delta_2$ . Note that

$$\frac{1}{nT_1} \frac{\partial \log |\Omega^*(\delta)|}{\partial \delta_{(j)}} = \frac{1}{nT_1} \text{tr} \left( \Omega^*(\delta)^{-1} \frac{\partial \Omega^*(\delta)}{\partial \delta_{(j)}} \right)$$

where  $\delta_{(j)}$  denotes the  $j$ th element of  $\delta$ ,  $j = 1, 2, 3, 4$ . We can use the explicit expression of  $\Omega^*(\delta)$  in (3.11) and show that  $\frac{1}{nT_1} \frac{\partial \log |\Omega^*(\delta)|}{\partial \delta_{(j)}} = O(1)$  uniformly in  $\delta$  for each  $j$ . This implies that  $\log |\Omega^*(\delta)|$  is uniformly equicontinuous in  $\delta$  on  $\Delta$ . As in the proof of Theorem 4.1, we can readily verify by contradiction that  $\tilde{\sigma}_v^2(\delta)$  is uniformly bounded away from zero on  $\Delta$ , and prove that  $\log [\tilde{\sigma}_v^2(\delta)]$  is uniformly equicontinuous on  $\Delta$  by showing that  $\tilde{\sigma}_v^2(\delta)$  is uniformly equicontinuous on  $\Delta$ . By (C.17), we have

$$\begin{aligned} \tilde{\sigma}_v^2(\delta) &= \frac{1}{nT_1} E(u^{*'} \Omega^{*-1} u^*) + \frac{1}{nT_1} \theta^*(\delta)' E(X^{*'} \Omega^{*-1} X^*) \theta^*(\delta) + \frac{(\rho_0 - \rho)^2}{nT_1} E(Y_{-1}^{*'} \Omega^{*-1} Y_{-1}^*) \\ &\quad + \frac{2}{nT_1} \theta^*(\delta)' E(X^{*'} \Omega^{*-1} u^*) + \frac{2(\rho_0 - \rho)}{nT_1} E(Y_{-1}^{*'} \Omega^{*-1} u^*) + \frac{2}{nT_1} \theta^*(\delta)' E(X^{*'} \Omega^{*-1} Y_{-1}^*). \end{aligned}$$

We can show the uniform equicontinuity of  $\tilde{\sigma}_v^2(\delta)$  by showing that of each of the six terms on the right hand side of the last equation. Using  $\tilde{\sigma}_v^2(\delta_0) = \sigma_{v0}^2$  again, we have

$$\begin{aligned} \frac{1}{nT_1} [\mathcal{L}_c^{rr*}(\delta) - \mathcal{L}_c^{rr*}(\delta_0)] &= \frac{1}{2nT_1} \{ \log |\Omega_0^*| - \log |\Omega^*(\delta)| \} + \frac{1}{2} \{ \log [\tilde{\sigma}_v^2(\delta_0)] - \log [\tilde{\sigma}_v^2(\delta)] \} \\ &= \frac{1}{2nT_1} \{ \log |\sigma_{v0}^2 \Omega_0^*| - \log [\tilde{\sigma}_v^2(\delta) \Omega^*(\delta)] \}. \end{aligned}$$

We can show that the identification uniqueness condition holds by using the uniform equicontinuity of  $\mathcal{L}_c^{rr*}(\delta)$  and a counter argument under Assumption R( $iv^*$ ). ■

**Proof of Theorem 4.4.** The proof is analogous to that of Theorem 4.2, but follows mainly from Lemmas B.12-B.13. ■

**Proof of Theorem 4.5.** As in the proof of Theorem 4.1, we prove the theorem by showing that (i)  $\frac{1}{nT_1} [\mathcal{L}_c^{f*}(\delta) - \mathcal{L}_c^f(\delta)] \xrightarrow{p} 0$  uniformly in  $\delta \in \Delta$ , and (ii)  $\limsup_{n \rightarrow \infty} \max_{\delta \in N_\epsilon^c(\delta_0)} \frac{1}{nT} [\mathcal{L}_c^{f*}(\delta) - \mathcal{L}_c^{f*}(\delta_0)] < 0$  for any  $\epsilon > 0$ .

By (3.26) and (4.9),  $\frac{2}{nT_1} [\mathcal{L}_c^{f*}(\delta) - \mathcal{L}_c^f(\delta)] = \log \hat{\sigma}_v^2(\delta) - \log \tilde{\sigma}_v^2(\delta)$ . To show (i), it suffices to show

$$\hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta) = o_p(1) \text{ uniformly on } \Delta \quad (\text{C.20})$$

provided  $\tilde{\sigma}_v^2(\delta)$  is uniformly bounded away from 0. By the definition of  $\tilde{\Delta}u(\delta)$  below (3.25), we have  $\tilde{\Delta}u(\delta) = \Delta Y^\dagger(\rho) - \Delta X^\dagger(\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta X^\dagger)^{-1}\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta Y^\dagger(\rho) = \Omega^{\dagger 1/2}M^\dagger\Omega^{\dagger-1/2}\Delta Y^\dagger(\rho)$  where  $M^\dagger = I_{nT} - \Omega^{\dagger-1/2}\Delta X^\dagger(\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta X^\dagger)^{-1}\Delta X^{\dagger'}\Omega^{\dagger-1/2}$  is a projection matrix. Observe that  $\Delta Y^\dagger(\rho) = \Delta Y^\dagger(\rho_0) + [\Delta Y^\dagger(\rho) - \Delta Y^\dagger(\rho_0)] = \Delta X^\dagger\theta_0 + \Delta u + (\rho_0 - \rho)\Delta Y_{-1}^\dagger$  where  $\Delta Y_{-1}^\dagger = (0_{1 \times n}, \Delta y'_1, \dots, \Delta y'_{T-1})'$ . This, in conjunction with the fact that  $M^\dagger\Omega^{\dagger-1/2}\Delta X^\dagger = 0$ , implies that

$$\begin{aligned}\hat{\sigma}_v^2(\delta) &= \frac{1}{nT_1}\tilde{\Delta}u(\delta)'\Omega^{\dagger-1}\tilde{\Delta}u(\delta) = \frac{1}{nT_1}\Delta Y^\dagger(\rho)'\Omega^{\dagger-1/2}M^\dagger\Omega^{\dagger-1/2}\Delta Y^\dagger(\rho) \\ &= \frac{1}{nT_1}[\Delta u + (\rho_0 - \rho)\Delta Y_{-1}^\dagger]'\Omega^{\dagger-1/2}M^\dagger\Omega^{\dagger-1/2}[\Delta u + (\rho_0 - \rho)\Delta Y_{-1}^\dagger].\end{aligned}\quad (C.21)$$

By (4.7) and the above expression for  $\Delta Y^\dagger(\rho)$ , we have  $\tilde{\theta}(\delta) = [E(\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta X^\dagger)]^{-1}E[\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta Y^\dagger(\rho)] = \theta_0 - \theta^\dagger(\delta)$ , where  $\theta^\dagger(\delta) = (\rho - \rho_0)[E(\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta X^\dagger)]^{-1}E(\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta Y_{-1}^\dagger)$ . Then by the definition of  $\Delta u(\theta, \rho)$  after (3.24),  $\Delta u(\tilde{\theta}(\delta), \rho) = \Delta Y^\dagger(\rho) - \Delta X^\dagger\tilde{\theta}(\delta) = \Delta X^\dagger\theta^\dagger(\delta) + \Delta u + (\rho_0 - \rho)\Delta Y_{-1}^\dagger$ . By (4.8),

$$\tilde{\sigma}_v^2(\delta) = \frac{1}{nT}E[v^\dagger(\delta)] + \frac{(\rho_0 - \rho)}{nT}\theta^\dagger(\delta)'E(\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta Y_{-1}^\dagger). \quad (C.22)$$

where  $v^\dagger(\delta) = [\Delta u + (\rho_0 - \rho)\Delta Y_{-1}^\dagger]'\Omega^{\dagger-1}[\Delta u + (\rho_0 - \rho)\Delta Y_{-1}^\dagger]$ . Using (C.21), (C.22), and  $\Omega^{\dagger-1/2}M^\dagger\Omega^{\dagger-1/2} = \Omega^{\dagger-1} - \Omega^{\dagger-1}\Delta X^\dagger(\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta X^\dagger)^{-1}\Delta X^{\dagger'}\Omega^{\dagger-1}$ , we have

$$\begin{aligned}\hat{\sigma}_v^2(\delta) - \tilde{\sigma}_v^2(\delta) &= \frac{1}{nT}\{v^\dagger(\delta) - E[v^\dagger(\delta)]\} + Q_{xu}^\dagger(\delta)'Q_{xx}^\dagger(\delta)^{-1}Q_{xu}^\dagger(\delta)' + 2(\rho_0 - \rho)Q_{xu}^\dagger(\delta)'Q_{xx}^\dagger(\delta)^{-1}Q_{xy-1}^\dagger(\delta) \\ &\quad + (\rho_0 - \rho)^2\{Q_{xy-1}^\dagger(\delta)'Q_{xx}^\dagger(\delta)^{-1}Q_{xy-1}^\dagger(\delta) - E[Q_{xy-1}^\dagger(\delta)']\{E[Q_{xx}^\dagger(\delta)]\}^{-1}E[Q_{xy-1}^\dagger(\delta)]\} \\ &\equiv \Pi_{n1}^\dagger(\delta) + \Pi_{n2}^\dagger(\delta) + 2(\rho_0 - \rho)\Pi_{n3}^\dagger(\delta) + (\rho_0 - \rho)^2\Pi_{n4}^\dagger(\delta), \text{ say,}\end{aligned}$$

where  $Q_{xx}^\dagger(\delta) = \frac{1}{nT}\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta X^\dagger$ ,  $Q_{xu}^\dagger(\delta) = \frac{1}{nT}\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta u$ , and  $Q_{xy-1}^\dagger(\delta) = \frac{1}{nT}\Delta X^{\dagger'}\Omega^{\dagger-1}\Delta Y_{-1}^\dagger$ . We prove (C.20) by showing that  $\Pi_{ns}^\dagger(\delta) = o_p(1)$  uniformly in  $\delta$  for  $s = 1, 2, 3$ , and 4. Analogously to the analysis of  $\Pi_{n1}^*(\delta)$  in the proof of Theorem 4.3, we can show that  $\Pi_{n1}^\dagger(\delta) = o_p(1)$  uniformly in  $\delta$ . By Assumptions G2 and F(iii),

$$\begin{aligned}|\Pi_{n2}^\dagger(\delta)| &\leq [\inf_{\delta \in \Delta} \lambda_{\min}(Q_{xx}^\dagger(\delta))]^{-1}\|Q_{xu}^\dagger(\delta)\|^2 \leq [\lambda_{\max}(\Omega^\dagger)]^{-1}\lambda_{\min}(\frac{1}{nT}\Delta X^{\dagger'}\Delta X^\dagger)\|Q_{xu}^\dagger(\delta)\|^2 \\ &= O(1)O_p(1)o_p(1) = o_p(1) \text{ uniformly in } \delta\end{aligned}$$

as we can readily show that  $Q_{xu}^\dagger(\delta) = o_p(1)$  uniformly in  $\delta$ . For  $\Pi_{n3}^\dagger(\delta)$ , we have

$$\Pi_{n3}^\dagger(\delta) \leq \|Q_{xu}^\dagger(\delta)\|\|Q_{xx}^\dagger(\delta)^{-1}\|\|Q_{xy-1}^\dagger(\delta)\| = o_p(1) \text{ uniformly in } \delta$$

as one can readily show that  $Q_{xy-1}^\dagger(\delta) = O_p(1)$  uniformly in  $\delta$ . The analysis of  $\Pi_{n4}^\dagger(\delta)$  is analogous to that of  $\Pi_{n4}^*(\delta)$ . This completes the proof of (i).

The proof of (ii) is analogous to that of part (ii) in the proofs of Theorem 4.1 and 4.3 and we only sketch the major differences. First, by the use of an auxiliary process, Jensen inequality, and the fact that  $\tilde{\sigma}_v^2(\delta_0) = \frac{1}{nT}E[\Delta u'\Omega^{\dagger-1}\Delta u] = \sigma_{v0}^2$  by (C.22), we can show that

$$\mathcal{L}_c^{f*}(\delta) \leq \mathcal{L}_c^{f*}(\delta_0) \text{ for any } \delta = (\lambda, \rho, \phi_\zeta)' \in \Delta.$$

As before, we show that  $\frac{1}{nT}\mathcal{L}_c^{f*}(\delta)$  is uniformly equicontinuous on  $\Delta$  by showing the uniform equicontinuity of  $\frac{1}{nT}\log|\Omega^\dagger(\delta)|$  and  $\log[\tilde{\sigma}_v^2(\delta)]$  on  $\Delta$ . Noting that  $\frac{1}{nT}\frac{\partial \log|\Omega^\dagger(\delta)|}{\partial \delta_{(j)}} = \frac{1}{nT}\text{tr}(\Omega^\dagger(\delta)^{-1}\frac{\partial \Omega^\dagger(\delta)}{\partial \delta_{(j)}})$  where

$\delta_{(j)}$  denotes the  $j$ th element of  $\delta$ ,  $j = 1, 2, 3$ , we can use the explicit expression of  $\Omega^\dagger(\delta)$  in (3.23) and show that  $\frac{1}{nT} \frac{\partial \log |\Omega^\dagger(\delta)|}{\partial \delta_{(j)}} = O(1)$  uniformly in  $\delta$  for each  $j$ . This implies that  $\log |\Omega^\dagger(\delta)|$  is uniformly equicontinuous in  $\delta$  on  $\Delta$ . As in the proof of Theorem 4.1, we can readily verify by contradiction that  $\tilde{\sigma}_v^2(\delta)$  is uniformly bounded away from zero on  $\Delta$ , and prove that  $\log[\tilde{\sigma}_v^2(\delta)]$  is uniformly equicontinuous on  $\Delta$  by showing that  $\tilde{\sigma}_v^2(\delta)$  is uniformly equicontinuous on  $\Delta$ . Now, by (4.8) and (C.22) we have

$$\begin{aligned} \tilde{\sigma}_v^2(\delta) &= \frac{1}{nT} E(\Delta u \Omega^{\dagger-1} \Delta u) + \frac{1}{nT} \theta^\dagger(\delta)' E(\Delta X^{\dagger'} \Omega^{\dagger-1} \Delta X^\dagger) \theta^\dagger(\delta) + \frac{(\rho_0 - \rho)^2}{nT} E(\Delta Y_{-1}^{\dagger'} \Omega^{\dagger-1} \Delta Y_{-1}^\dagger) \\ &\quad + \frac{2}{nT} \theta^\dagger(\delta)' E(\Delta X^{\dagger'} \Omega^{\dagger-1} \Delta u) + \frac{2(\rho_0 - \rho)}{nT} E(\Delta Y_{-1}^{\dagger'} \Omega^{\dagger-1} \Delta u) + \frac{2}{nT} \theta^\dagger(\delta)' E(\Delta X^{\dagger'} \Omega^{\dagger-1} \Delta Y_{-1}^\dagger). \end{aligned}$$

We can show the uniform equicontinuity of  $\tilde{\sigma}_v^2(\delta)$  by showing that of each of the six terms on the right hand side of the last equation. Using  $\tilde{\sigma}_v^2(\delta_0) = \sigma_{v0}^2$  again, we have

$$\begin{aligned} \frac{1}{nT} [\mathcal{L}_c^{f*}(\delta) - \mathcal{L}_c^{f*}(\delta_0)] &= \frac{1}{2nT} \{ \log |\Omega_0^\dagger| - \log |\Omega^\dagger(\delta)| \} + \frac{1}{2} \{ \log[\tilde{\sigma}_v^2(\delta_0)] - \log[\tilde{\sigma}_v^2(\delta)] \} \\ &= \frac{1}{2nT} \{ \log |\sigma_{v0}^2 \Omega_0^\dagger| - \log |\tilde{\sigma}_v^2(\delta) \Omega^\dagger(\delta)| \}. \end{aligned}$$

Then we can show the identification uniqueness condition by using the uniform equicontinuity of  $\mathcal{L}_c^{f*}(\delta)$  and a counter argument under Assumption F. ■

**Proof of Theorem 4.6.** The proof is analogous to that of Theorem 4.2, but follows mainly from Lemmas B.14-B.15. ■

## Appendix D: Proofs of the Propositions in Section 5

### Proof of Proposition 5.1

Decompose the score component for  $\theta_0$  according to  $(\beta'_0, \gamma'_0)'$  and  $\rho_0$ :  $\sigma_{v0}^{-2} \mathbf{X} \Omega_0^{-1} u$  and  $\sigma_{v0}^{-2} Y'_{-1} \Omega_0^{-1} u$ , where  $\mathbf{X} = (X, Z)$ . Write (B.3) as  $Y_{-1} = \boldsymbol{\eta}_{-1} + (\mathcal{J}'_{\rho_0} \otimes I_n)u$ , where  $\boldsymbol{\eta}_{-1}$  is the exogenous part of  $Y_{-1}$  and  $\mathcal{J}_{\rho_0}$  is given in (B.4). The score vector  $S^r(\psi_0)$  is thus expressed in terms of  $\psi_0$  and  $u$ ,

$$S^r(\psi_0) = \begin{cases} \frac{1}{\sigma_{v0}^2} \mathbf{X}' \Omega_0^{-1} u \\ \frac{1}{\sigma_{v0}^2} \boldsymbol{\eta}'_{-1} \Omega_0^{-1} u + \frac{1}{\sigma_{v0}^2} u' (\mathcal{J}_{\rho_0} \otimes I_n) \Omega_0^{-1} u \\ \frac{1}{2\sigma_{v0}^2} u' P_\omega u - \frac{1}{2} \text{tr}(P_\omega \Omega_0), \quad \text{for } \omega = \sigma_v^2, \lambda_0, \phi_{\mu 0}, \end{cases}$$

where  $P_\omega = \frac{1}{\sigma_{v0}^2} \Omega_0^{-1}$ ,  $\Omega_0^{-1} (I_T \otimes A_0) \Omega_0^{-1}$ , and  $\Omega_0^{-1} (J_T \otimes I_n) \Omega_0^{-1}$ , or in terms of  $\psi_0$  and  $\mathbf{r}$  through  $\mathbf{r} \equiv (r'_1, \dots, r'_T)' = (I_T \otimes \Sigma_0^{-\frac{1}{2}})u$  where  $\Sigma_0 = \Sigma(\lambda_0, \phi_{\mu 0})$ . The score vector is seen to contain three types of terms: quadratic  $\mathbf{r}' Q \mathbf{r}$ , linear  $R' \mathbf{r}$  and constant  $C$ . The result follows if  $\frac{1}{nT} [\text{Var}^b(\hat{R}' \hat{\mathbf{r}}^b) - \text{Var}(R' \mathbf{r})] = o_p(1)$ ,  $\frac{1}{nT} [\text{Var}^b(\hat{\mathbf{r}}^b \hat{Q} \hat{\mathbf{r}}^b) - \text{Var}(\mathbf{r}' Q \mathbf{r})] = o_p(1)$ , and  $\frac{1}{nT} [\text{Cov}^b(\hat{R}' \hat{\mathbf{r}}^b, \hat{\mathbf{r}}^b \hat{Q} \hat{\mathbf{r}}^b) - \text{Cov}(R' \mathbf{r}, \mathbf{r}' Q \mathbf{r})] = o_p(1)$ , where  $\hat{R}$  and  $\hat{Q}$  are the QMLEs of  $R$  and  $Q$ . Similarly,  $\hat{\Sigma}$  and  $\hat{S}$  used latter are the QMLEs of  $\Sigma_0$  and  $S$ .

Without loss of generality, let  $T = 2$ . Thus,  $\mathbf{r} = (r'_1, r'_2)'$  and  $\hat{\mathbf{r}} = (\hat{r}'_1, \hat{r}'_2)'$ . Note that  $\hat{r}_1^b$  and  $\hat{r}_2^b$  are two matched bootstrap samples, corresponding to  $n$  random draws from the rows of  $\{\hat{r}_1, \hat{r}_2\}$ . Note also that  $r_t = \Sigma_0^{-\frac{1}{2}} u_t = \Sigma_0^{-\frac{1}{2}} \mu + \Sigma_0^{-\frac{1}{2}} B_0^{-1} v_t$  and the matrices  $\Sigma_0$  and  $B_0$  depend mainly on the spatial weight matrix  $W$ . Let  $\omega_{ij}$  and  $\omega_{ij}^*$  be, respectively, the elements of  $\Sigma_0^{-\frac{1}{2}}$  and  $\Sigma_0^{-\frac{1}{2}} B_0^{-1}$ , and let  $\varpi_{ii}$  be the diagonal elements of  $\Sigma_0^{-1}$ . We consider standard  $W$  matrices so that the following results maintain: (i) for  $k \geq 3$ ,  $\sum_{i,j} \omega_{ij}^k \sim \sum_i \omega_{ii}^k$  and  $\sum_{i,j} \omega_{ij}^k \sim \sum_i \omega_{ii}^{*k}$ , and (ii) for  $k \geq 3$ ,  $\mathbf{s}^2(\omega_{ii}^k) = o(1)$  and  $\mathbf{s}^2(\omega_{ii}^{*k}) = o(1)$ , and (iii)  $\mathbf{s}^2(\varpi_{ii}) = o(1)$ , where, e.g.,  $\mathbf{s}^2(\omega_{ii}^k)$  denotes the sample variance of  $\{\omega_{ii}^k\}$ .

Letting  $R = (R'_1, R'_2)'$ , we have for the linear terms,  $\text{Var}(R'\mathbf{r}) = \sigma_{v0}^2(R'_1R_1 + R'_2R_2) + 2\sigma_{\mu0}^2R'_1\Sigma_0^{-1}R_2$ , and  $\text{Var}^b(\hat{R}'\hat{\mathbf{r}}^b) = \hat{\sigma}_v^2(\hat{R}'_1\hat{R}_1 + \hat{R}'_2\hat{R}_2) + \frac{2}{n}\hat{\sigma}_\mu^2\text{tr}(\hat{\Sigma}^{-1})\hat{R}'_1\hat{R}_2$ . It follows that

$$\frac{1}{nT}[\text{Var}^b(\hat{R}'\hat{\mathbf{r}}^b) - \text{Var}(R'\mathbf{r})] = -\frac{2}{nT}\sigma_{\mu0}^2R'_1[\Sigma_0^{-1} - \frac{1}{n}\text{tr}(\Sigma_0^{-1})I_n]R_2 + o_p(1) = o_p(1),$$

for  $R' = \frac{1}{\sigma_{v0}^2}\mathbf{X}'\Omega_0^{-1}(I_T \otimes \Sigma_0^{\frac{1}{2}})$ , or  $\frac{1}{\sigma_{v0}^2}\boldsymbol{\eta}'_{-1}\Omega_0^{-1}(I_T \otimes \Sigma_0^{\frac{1}{2}})$ , by Assumption G1(iii) and the result (iii) above.

For the quadratic terms, partitioning  $Q$  as  $\{Q_{ts}\}$  according to  $t, s = 1, 2$ , we have,  $\text{Var}(\mathbf{r}'Q\mathbf{r}) = \text{Var}(r'_1Q_{11}r_1 + r'_2Q_{22}r_2 + r'_1Q_{12}r_2 + r'_2Q_{21}r_1) = \text{Var}(r'_1Q_{11}r_1) + \dots + 2\text{Cov}(r'_1Q_{11}r_1, r'_2Q_{22}r_2) + \dots$ , and similarly,  $\text{Var}^b(\hat{\mathbf{r}}^b'\hat{Q}\hat{\mathbf{r}}^b) = \text{Var}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1) + \dots + 2\text{Cov}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1, \hat{r}^{b'}_2\hat{Q}_{22}\hat{r}^b_2) + \dots$ . It boils down to show that  $\frac{1}{n}[\text{Var}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1) - \text{Var}(r'_1Q_{11}r_1)] = o_p(1)$ ,  $\frac{1}{n}[\text{Cov}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1, \hat{r}^{b'}_2\hat{Q}_{22}\hat{r}^b_2) - \text{Cov}(r'_1Q_{11}r_1, r'_2Q_{22}r_2)] = o_p(1)$ , etc. We formally prove these two terms, and others follow in a similar fashion. It is easy to show that

$$\begin{aligned}\text{Var}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1) &= \hat{\kappa}_{r_1}\hat{q}'_{11}\hat{q}_{11} + \hat{\sigma}_{v_1}^4\text{tr}[\hat{\Sigma}\hat{S}_{11}\hat{\Sigma}(\hat{S}_{11} + \hat{S}'_{11})], \text{ and} \\ \text{Var}(r'_1Q_{11}r_1) &= \kappa_\mu s'_{11}s_{11} + \kappa_v s'^*_{11}s^*_{11} + \sigma_{v0}^4\text{tr}[\Sigma_0S_{11}\Sigma_0(S_{11} + S'_{11})],\end{aligned}$$

where  $S_{11} = \Sigma_0^{-\frac{1}{2}}Q_{11}\Sigma_0^{-\frac{1}{2}}$ ,  $s_{11} = \text{diagv}(S_{11})$ ,  $s^*_{11} = \text{diagv}(B_0^{-1}S_{11}B_0^{-1})$ ,  $\hat{q}_{11} = \text{diagv}(\hat{Q}_{11})$ , and  $\hat{\sigma}_{v_1}^2$  and  $\hat{\kappa}_{r_1}$  are the 2nd and 4th sample cumulants of  $\hat{r}_1$ . It follows that

$$\frac{1}{n}[\text{Var}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1) - \text{Var}(u'_1S_{11}u_1)] = \frac{1}{n}[\hat{\kappa}_r\hat{q}'_{11}\hat{q}_{11} - \kappa_\mu s'_{11}s_{11} - \kappa_v s'^*_{11}s^*_{11}] + o_p(1)$$

Furthermore,  $\frac{1}{n}\hat{q}'_{11}\hat{q}_{11} = \frac{1}{n}q'_{11}q_{11} + o_p(1)$ , and the results (i)-(iii) above lead to the following:

$$\begin{aligned}\hat{\kappa}_r &= \frac{1}{n}\sum_{i=1}^n\hat{r}_i^4 - 3\hat{\sigma}_{v_1}^4 = \frac{\kappa_\mu}{n}\sum_{i,j}\omega_{ij}^4 + \frac{\kappa_v}{n}\sum_{i,j}\omega_{ij}^{*4} + o_p(1), \\ \frac{1}{n^2}q'_{11}q_{11}\sum_{i,j}\omega_{ij}^4 - \frac{1}{n}s'_{11}s_{11} &= \frac{1}{n^2}q'_{11}q_{11}\sum_i\omega_{ii}^4 - \frac{1}{n}\sum_iq_{11,ii}^2\omega_{ii}^4 + o(1) = o(1), \text{ and} \\ \frac{1}{n^2}q'_{11}q_{11}\sum_{i,j}\omega_{ij}^{*4} - \frac{1}{n}s'^*_{11}s^*_{11} &= \frac{1}{n^2}q'_{11}q_{11}\sum_i\omega_{ii}^{*4} - \frac{1}{n}\sum_iq_{11,ii}^2\omega_{ii}^{*4} + o(1) = o(1).\end{aligned}$$

It follows that  $\frac{1}{n}[\text{Var}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1) - \text{Var}(u'_1S_{11}u_1)] = o_p(1)$ .

Now,  $\text{Cov}(r'_1Q_{11}r_1, r'_2Q_{22}r_2) = \kappa_\mu s'_{11}s_{22} + \sigma_{\mu0}^4\text{tr}[S_{11}(S_{22} + S'_{22})]$ , and  $\text{Cov}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1, \hat{r}^{b'}_2\hat{Q}_{22}\hat{r}^b_2) = (\hat{\rho}_2 - \hat{\sigma}_v^4 - 2\hat{\rho}_1^2)\hat{q}'_{11}\hat{q}_{22} + \hat{\rho}_1^2\text{tr}[\hat{Q}_{11}(\hat{Q}_{22} + \hat{Q}'_{22})]$ , where  $\hat{\rho}_1 = \text{E}^b(\hat{r}^b_{1i}\hat{r}^b_{2i}) = \sigma_{\mu0}^2(\frac{1}{n}\sum_i\varpi_{ii}) + o_p(1)$ , and  $\hat{\rho}_2 = \text{E}^b((\hat{r}^b_{1i}\hat{r}^b_{2i})^2) = \kappa_\mu(\frac{1}{n}\sum_{i,j}\omega_{ij}^4) + \sigma_{v0}^4 + 2\sigma_{\mu0}^4(\frac{1}{n}\sum_i\varpi_{ii}^2) + o_p(1)$ . Thus, by the results (i)-(iii) above,

$$\begin{aligned}&\frac{1}{n}[\text{Cov}^b(\hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1, \hat{r}^{b'}_2\hat{Q}_{22}\hat{r}^b_2) - \text{Cov}(u'_1S_{11}u_1, u'_2S_{22}u_2)] \\ &= \frac{\kappa_\mu}{n}(q'_{11}q_{22}\frac{1}{n}\sum_{i,j}\omega_{ij}^4 - s'_{11}s_{22}) + \frac{\sigma_{\mu0}^4}{n}\text{tr}[(\bar{\omega}^2Q_{11} - \Sigma_0^{-1}Q_{11}\Sigma_0^{-1})(Q_{22} + Q'_{22})] + o_p(1) = o_p(1),\end{aligned}$$

where  $\bar{\omega} = \frac{1}{n}\sum_i\varpi_{ii}$ .

Finally, the proof of  $\frac{1}{nT}[\text{Cov}^b(\hat{R}'\hat{\mathbf{r}}^b, \hat{\mathbf{r}}^{b'}\hat{Q}\hat{\mathbf{r}}^b) - \text{Cov}(R'\mathbf{r}, \mathbf{r}'Q\mathbf{r})] = o_p(1)$  can be carried out in the same manner but is simpler. We give detail below for the most complicate term, and others follow.

$$\text{Cov}^b(\hat{R}'_1\hat{r}^b_1, \hat{r}^{b'}_1\hat{Q}_{11}\hat{r}^b_1) - \text{Cov}(R'_1r_1, r'_1Q_{11}r_1) = \hat{\gamma}_{r_1}\hat{R}'_1\hat{q}_{11} - R'_1\Sigma_0^{-\frac{1}{2}}(\gamma_\mu s_{11} + \gamma_v B_0^{-1}s^*_{11}),$$

where  $\hat{\gamma}_{r_1}$  is the 3rd sample cumulant of  $\hat{r}_1$ , and  $\gamma_\mu$  and  $\gamma_v$  are the 3rd cumulants of  $\mu_i$  and  $v_{it}$ , respectively. It is easy to show that  $\hat{\gamma}_{r_1} = \frac{1}{n}\sum_i\hat{r}_i^3 = \frac{1}{n}\sum_i r_i^3 + o_p(1) = \gamma_\mu(\frac{1}{n}\sum_{i,j}\omega_{ij}^3) + \gamma_v(\frac{1}{n}\sum_{i,j}\omega_{ij}^{*3})$ , and that,

$$\frac{1}{n}[\hat{R}'_1\hat{q}_{11}(\frac{1}{n}\sum_{i,j}\omega_{ij}^3) - R'_1\Sigma_0^{-\frac{1}{2}}s_{11}] = o_p(1), \text{ and } \frac{1}{n}[\hat{R}'_1\hat{q}_{11}(\frac{1}{n}\sum_{i,j}\omega_{ij}^{*3}) - R'_1\Sigma_0^{-\frac{1}{2}}B_0^{-1}s^*_{11}] = o_p(1).$$

The result thus follows. ■



**Proof of Proposition 5.2.** From (3.10),  $y_0 = \eta_0 + u_0$  where  $\eta_0$  denotes the exogenous part of  $y_0$ . The key element  $u_\rho^* = -(\dot{a}_m \gamma' Z', Y_{-1}')'$  in the score function  $S^{rr}(\psi_0)$  can be expressed as  $-\eta_{-1}^* - (\mathcal{J}_\rho^{*'} \otimes I_n)u^*$ , where  $\mathcal{J}_\rho^*$  extends  $\mathcal{J}_\rho$  by adding a column  $(\rho^{T-1}, \rho^{T-2}, \dots, 1, 0)'$  on its right and a row of zeros at its bottom, and  $-\eta_{-1}^*$  is the exogenous part of  $u_\rho^*(\theta_0, \rho_0)$ . Thus,  $S^{rr}(\psi)$  is expressed in terms of  $\psi_0$ , and linear and quadratic forms of  $u^*$ . The proof proceeds as that of Proposition 5.1. ■

**Proof of Proposition 5.3.** From (3.19),  $\Delta y_1 = \Delta \eta_1 + \Delta \tilde{u}_1$  where  $\Delta \eta_1$  denotes the exogenous part of  $\Delta y_1$ . The key element  $\Delta u_\rho = -(0'_{n \times 1}, \Delta y'_1, \dots, \Delta y'_{T-1})'$  in the score function  $S^f(\psi_0)$  can be expressed as  $-\Delta \eta_{-1} - (\mathcal{J}_\rho' \otimes I_n)\Delta u$ , where  $-\Delta \eta_{-1}$  denotes the exogenous part of  $\Delta u_\rho$ . Subsequently,  $S^f(\psi_0)$  is expressed in terms of  $\psi_0$ , and linear and quadratic forms of  $\Delta u$ . The proof proceeds as that of Proposition 5.1. ■

## REFERENCES

- Anderson, T. W., Hsiao, C., 1981. Estimation of dynamic models with error components. *Journal of American Statistical Association* 76, 598-606.
- Anderson, T. W., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18, 47-82.
- Andrews, D. W. K., 2005. Cross-section regression with common shocks. *Econometrica* 73, 1551-1585.
- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. The Netherlands: Kluwer Academic Press.
- Anselin, L., 2001. Spatial econometrics. In: Baltagi, B. H. (Eds.), *A companion to theoretical econometrics*. Blackwell Publishers Ltd., Massachusetts, pp. 310-330.
- Anselin, L., Le Gallo, J., Jayet, J., 2008. Spatial panel econometrics. In: Mátyás, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Springer-Verlag, Berlin Heidelberg, pp. 625-660.
- Bai, J., 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 1229-1279.
- Balestra, P., Nerlove, M., 1966. Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas. *Econometrica* 34, 585-612.
- Baltagi, B. H., Li, D., 2004. Prediction in the panel data models with spatial correlation. In: Anselin, L., Florax, R., Rey, S. J. (Eds.), *Advances in Spatial Econometrics*. Springer-Verlag, Berlin, pp. 283-296.
- Baltagi, B. H., Song, S. H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117, 123-150.
- Baltagi, B. H., Song, S. K., Jung, B. C., Koh, W., 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140, 5-51.
- Baltagi, B. H., Egger, P., Pfaffermayr, M., 2013. A generalized spatial panel model with random effects. *Econometric Reviews* 32, 650-685.
- Bhargava, A., Sargan, J. D., 1983. Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51, 1635-1659.
- Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton.
- Binder, M., Hsiao, C., Pesaran, M. H., 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795-837.
- Blundell, R., Smith, R. J., 1991. Conditions initiales et estimation efficace dans les modèles dynamiques sur données de panel. *Annales d'Économie et de Statistique* 20/21, 109-123.

- Chen, X., Conley, T. G., 2001. A new semiparametric spatial model for panel time series. *Journal of Econometrics* 105, 59-83.
- Cliff, A. D., Ord, J. K., 1973. *Spatial Autocorrelation*. Pion Ltd., London.
- Davidson, J., 1994. *Stochastic Limit Theory*. Oxford University Press, Oxford.
- Elhorst, J. P., 2003. Specification and estimation of spatial panel data models. *International Regional Science Review* 26, 244-268.
- Elhorst, J. P., 2005. Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37, 85-106.
- Elhorst, J. P., 2010. Dynamic panels with endogenous interaction effects when  $T$  is small. *Regional Science and Urban Economics* 40, 272-282.
- Griffith, D. A., 1988. *Advanced Spatial Statistics*. Kluwer, Dordrecht, the Netherlands.
- Horn, R., Johnson, C., 1985. *Matrix Analysis*. John Wiley & Sons, New York.
- Hsiao, C., Pesaran, M. H., Tahmiscioglu, A. K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107-150.
- Hsiao, C., 2003. *Analysis of Panel Data*. 2nd edition. Cambridge University Press, Cambridge.
- Kapoor, M., Kelejian, H. H., Prucha, I. R., 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97-130.
- Kelejian, H. H., Prucha, I. R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17, 377-398.
- Kelejian, H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.
- Kelejian, H. H., Prucha, I. R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219-257.
- Kelejian, H. H., Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53-67.
- Lee, L. F., 2002. *Asymptotic Distribution of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models*, Mimeo, Dept. of Economics, Ohio State University.
- Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.
- Lee, L. F., Yu, J., 2010a. Estimation of spatial panel model with fixed effects. *Journal of Econometrics* 154, 165-185.
- Lee, L. F., Yu, J., 2010b. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26, 564-597.
- Lee, L. F., Yu, J., 2010c. Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40, 255-271.
- LeSage, J., Pace, R.K., 2009. *Introduction to Spatial Econometrics*. CRC Press, Taylor & Francis Group, London.
- Lin, X., Lee, L. F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157, 34-52.
- Maddala, G. S., 1971. The use of variance components models in pooling cross section and time series data. *Econometrica* 39, 341-358.
- Magnus, J. R., 1982. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *Journal of Econometrics* 19, 239-285.

- Magnus, J. R., Neudecker, H., 2002 Matrix Differential Calculus with Applications in Statistics and Econometrics. John Wiley and Sons, New York.
- Moon, H. & M. Weidner (2013) Dynamic linear panel regression models with interactive fixed effects. Manuscript. Working paper, University of Southern California.
- Mutl, J., 2006. Dynamic panel data models with spatially correlated disturbances. Ph.D. Thesis, University of Maryland, College Park.
- Mutl, J., Pfaffermayr, M., 2011. The Hausman test in a Cliff and Ord panel model. *Econometrics Journal* 14, 48-76.
- Nerlove, M., 1971. Further evidence on the estimation of dynamic economic relations from a time series of cross sections. *Econometrica* 39, 359-382.
- Nerlove, M., 2002. Likelihood inference for dynamic panel models. In: *Essays in Panel Data Econometrics*, M. Nerlove (Eds.). Cambridge University Press, pp. 307-348.
- Ord, J. K., 1975. Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association* 70, 120-126.
- Parent, O., LeSage, J. P., 2011. A space-time filter for panel data models containing random effects. *Computational Statistics and Data Analysis* 55, 475-490.
- Pesaran, M. H., 2004. General diagnostic tests for cross section dependence in panels. Working Paper No. 1229, University of Cambridge.
- Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74, 967-1012.
- Pesaran, M. H., Tosetti, E., 2011. Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182-202.
- Phillips, P. C. B., Sul, D., 2003. Dynamic panel estimation and homogeneity testing under cross-section dependence. *Econometrics Journal* 6, 217-259.
- Ridder, G., Wansbeek, T., 1990. Dynamic models for panel data. In: van der Ploeg, F. (Eds.), *Advanced Lectures in Quantitative Economics*. London: Academic Press, pp. 557-582.
- Robinson, P. M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics* 157, 6-17.
- Smirnov, O., Anselin, L., 2001. Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Computational Statistics and Data Analysis* 35, 301-319.
- Su, L., 2012. A semiparametric GMM estimation of spatial autoregressive models. *Journal of Econometrics* 167, 543-560.
- Su, L., Jin, S., 2010. Profile quasi-maximum likelihood estimation of spatial autoregressive models. *Journal of Econometrics* 157, 18-33.
- Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169, 34-47.
- Su, L., Yang, Z., 2007. QML estimation of dynamic panel data models with spatial errors. Working Paper, Singapore Management University.
- White, H., 1994. Estimation, Inference and Specification Analysis. *Econometric Society Monographs* No. 22, Cambridge University Press, Cambridge.
- Yang, Z., Li, C., Tse, Y. K., 2006. Functional form and spatial dependence in dynamic panels. *Economics Letters* 91, 138-145.
- Yu, J., de Jong, R., Lee, L. F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large. *Journal of Econometrics* 146, 118-134.

Yu, J., Lee, L. F., 2010. Estimation of unit root spatial dynamic panel data models. *Econometric Theory* 26, 1332-1362.

**Table 1a.** Monte Carlo Mean[RMSE] for the QMLEs, Random Effects Model with Normal Errors

$\psi$	true $m = 0$			true $m = 6$		
	$m = 0$	$m = 6$	$m = 200$	$m = 0$	$m = 6$	$m = 200$
	$n = 50, T = 3$					
5.0	5.0266[0.334]	4.9604[0.338]	5.0030[0.328]	4.5591[0.378]	4.9940[0.411]	5.0988[0.411]
1.0	1.0011[0.040]	0.9917[0.045]	0.9981[0.045]	0.9626[0.041]	0.9980[0.040]	1.0057[0.039]
1.0	0.9951[0.345]	0.9852[0.350]	0.9927[0.352]	0.7418[0.365]	0.9384[0.391]	0.9790[0.395]
<b>0.8</b>	0.7991[0.023]	0.8071[0.024]	0.8018[0.022]	0.8238[0.015]	0.8015[0.017]	0.7963[0.016]
0.5	0.4827[0.099]	0.3023[0.115]	0.2868[0.114]	0.4732[0.101]	0.4886[0.098]	0.4868[0.098]
1.0	0.9681[0.147]	0.1469[0.116]	0.0214[0.055]	0.8648[0.145]	0.9528[0.158]	0.9280[0.161]
1.0	0.9834[0.072]	1.2563[0.087]	1.2805[0.088]	1.0056[0.076]	0.9880[0.073]	1.0019[0.076]
5.0	4.9785[0.357]	4.9683[0.400]	4.9719[0.400]	4.7922[0.353]	5.0164[0.352]	5.0162[0.352]
1.0	1.0003[0.040]	0.9964[0.045]	0.9967[0.045]	0.9780[0.041]	0.9981[0.039]	0.9981[0.039]
1.0	0.9937[0.323]	1.0022[0.328]	1.0028[0.328]	0.8910[0.352]	0.9374[0.360]	0.9370[0.361]
<b>0.4</b>	0.4015[0.034]	0.4025[0.044]	0.4019[0.044]	0.4271[0.032]	0.4009[0.030]	0.4009[0.030]
0.5	0.4799[0.103]	0.3694[0.141]	0.3690[0.142]	0.4765[0.104]	0.4912[0.093]	0.4911[0.093]
1.0	0.9609[0.146]	0.6380[0.229]	0.6364[0.231]	0.9141[0.155]	0.9725[0.148]	0.9712[0.149]
1.0	0.9838[0.074]	1.1272[0.137]	1.1280[0.138]	1.0056[0.080]	0.9960[0.074]	0.9964[0.074]
5.0	5.0096[0.337]	4.9719[0.352]	4.9719[0.352]	4.9061[0.328]	5.0103[0.328]	5.0103[0.328]
1.0	0.9987[0.040]	0.9947[0.042]	0.9947[0.042]	0.9872[0.040]	0.9991[0.039]	0.9991[0.039]
1.0	0.9944[0.336]	0.9805[0.337]	0.9805[0.337]	0.9481[0.356]	0.9897[0.361]	0.9897[0.361]
<b>0.0</b>	-0.0014[0.041]	0.0069[0.047]	0.0069[0.047]	0.0199[0.043]	-0.0021[0.042]	-0.0021[0.042]
0.5	0.4783[0.106]	0.3977[0.114]	0.3977[0.114]	0.4815[0.102]	0.4929[0.091]	0.4929[0.091]
1.0	0.9659[0.151]	0.7313[0.178]	0.7313[0.178]	0.9342[0.157]	0.9691[0.148]	0.9691[0.148]
1.0	0.9808[0.076]	1.0741[0.102]	1.0741[0.102]	0.9945[0.079]	0.9624[0.066]	0.9624[0.066]
	$n = 100, T = 3$					
5.0	4.9921[0.252]	4.9129[0.258]	4.9423[0.248]	4.5604[0.270]	5.0174[0.299]	5.1460[0.300]
1.0	0.9995[0.029]	0.9892[0.034]	0.9932[0.033]	0.9655[0.029]	0.9997[0.029]	1.0090[0.029]
1.0	1.0019[0.243]	0.9822[0.242]	0.9916[0.242]	0.9112[0.227]	1.0126[0.240]	1.0414[0.244]
<b>0.8</b>	0.8003[0.017]	0.8092[0.018]	0.8058[0.016]	0.8200[0.009]	0.7993[0.010]	0.7935[0.010]
0.5	0.4852[0.074]	0.2674[0.086]	0.2500[0.085]	0.4857[0.068]	0.4872[0.067]	0.4865[0.067]
1.0	0.9788[0.101]	0.1828[0.094]	0.0279[0.056]	0.9083[0.101]	0.9806[0.115]	0.9719[0.120]
1.0	0.9941[0.052]	1.2885[0.062]	1.3150[0.060]	1.0075[0.053]	0.9940[0.052]	1.0025[0.053]
5.0	4.9941[0.247]	4.9271[0.305]	4.9318[0.306]	4.7258[0.277]	4.9982[0.273]	4.9982[0.273]
1.0	0.9991[0.031]	0.9899[0.040]	0.9904[0.040]	0.9730[0.031]	1.0012[0.030]	1.0012[0.030]
1.0	1.0055[0.242]	0.9888[0.245]	0.9897[0.245]	0.9384[0.240]	1.0127[0.250]	1.0128[0.250]
<b>0.4</b>	0.4004[0.025]	0.4104[0.037]	0.4098[0.037]	0.4316[0.023]	0.3996[0.022]	0.3996[0.022]
0.5	0.4916[0.069]	0.3706[0.099]	0.3701[0.100]	0.4885[0.074]	0.4859[0.069]	0.4858[0.069]
1.0	0.9885[0.103]	0.6050[0.175]	0.6033[0.177]	0.9141[0.104]	0.9808[0.101]	0.9798[0.101]
1.0	0.9926[0.053]	1.1742[0.118]	1.1752[0.118]	1.0120[0.054]	0.9948[0.051]	0.9951[0.051]
5.0	5.0098[0.265]	5.0200[0.271]	5.0200[0.271]	4.8775[0.257]	5.0054[0.254]	5.0054[0.254]
1.0	1.0011[0.032]	1.0023[0.033]	1.0023[0.033]	0.9845[0.032]	0.9997[0.030]	0.9997[0.030]
1.0	0.9923[0.232]	0.9930[0.233]	0.9930[0.233]	0.9819[0.240]	1.0086[0.244]	1.0086[0.244]
<b>0.0</b>	0.0000[0.031]	-0.0021[0.033]	-0.0021[0.033]	0.0236[0.033]	-0.0010[0.031]	-0.0010[0.031]
0.5	0.4860[0.069]	0.4257[0.073]	0.4258[0.073]	0.4866[0.072]	0.4942[0.063]	0.4942[0.063]
1.0	0.9771[0.107]	0.8260[0.117]	0.8261[0.117]	0.9505[0.109]	0.9851[0.101]	0.9851[0.101]
1.0	0.9957[0.054]	1.0535[0.068]	1.0535[0.068]	1.0015[0.054]	0.9778[0.045]	0.9778[0.045]

**Note:**  $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \sigma_\mu, \sigma_v)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.01, .5, .5, 2, 1)$  (see Footnote 13).

**Table 1b.** Monte Carlo Mean[RMSE] for the QMLEs, Random Effects Model with Normal Mixture

$\psi$	true $m = 0$			true $m = 6$		
	$m = 0$	$m = 6$	$m = 200$	$m = 0$	$m = 6$	$m = 200$
	$n = 50, T = 3$					
5.0	5.0194[0.342]	4.9734[0.350]	5.0140[0.340]	4.5754[0.416]	4.9935[0.429]	5.0941[0.430]
1.0	1.0005[0.039]	0.9948[0.047]	1.0006[0.047]	0.9656[0.041]	0.9984[0.039]	1.0057[0.039]
1.0	0.9874[0.335]	0.9778[0.339]	0.9858[0.340]	0.7650[0.383]	0.9558[0.405]	0.9981[0.410]
<b>0.8</b>	0.7992[0.022]	0.8047[0.024]	0.7998[0.022]	0.8225[0.017]	0.8011[0.016]	0.7960[0.016]
0.5	0.4788[0.100]	0.2652[0.130]	0.2489[0.129]	0.4766[0.099]	0.4916[0.097]	0.4902[0.096]
1.0	0.9544[0.249]	0.1551[0.120]	0.0283[0.061]	0.8470[0.228]	0.9330[0.259]	0.9101[0.260]
1.0	0.9792[0.145]	1.2519[0.163]	1.2776[0.167]	0.9984[0.147]	0.9821[0.143]	0.9954[0.147]
5.0	4.9914[0.340]	4.9151[0.373]	4.9190[0.374]	4.8085[0.368]	5.0216[0.361]	5.0215[0.361]
1.0	0.9990[0.042]	0.9887[0.047]	0.9891[0.047]	0.9814[0.040]	1.0002[0.038]	1.0002[0.038]
1.0	1.0152[0.332]	1.0061[0.333]	1.0067[0.333]	0.8921[0.357]	0.9384[0.361]	0.9381[0.361]
<b>0.4</b>	0.4003[0.033]	0.4120[0.041]	0.4114[0.041]	0.4265[0.033]	0.4016[0.030]	0.4016[0.030]
0.5	0.4784[0.099]	0.3775[0.115]	0.3770[0.116]	0.4804[0.097]	0.4914[0.090]	0.4913[0.090]
1.0	0.9488[0.256]	0.5328[0.299]	0.5307[0.302]	0.8779[0.250]	0.9387[0.249]	0.9375[0.249]
1.0	0.9799[0.144]	1.1476[0.183]	1.1485[0.184]	0.9895[0.148]	0.9770[0.138]	0.9774[0.138]
5.0	5.0179[0.343]	5.0602[0.344]	5.0602[0.344]	4.9083[0.343]	5.0085[0.339]	5.0085[0.339]
1.0	0.9990[0.044]	1.0016[0.044]	1.0016[0.044]	0.9884[0.040]	1.0000[0.038]	1.0000[0.038]
1.0	0.9981[0.343]	1.0043[0.344]	1.0043[0.344]	0.9497[0.346]	0.9928[0.349]	0.9929[0.349]
<b>0.0</b>	-0.0009[0.043]	-0.0094[0.043]	-0.0094[0.043]	0.0197[0.045]	-0.0017[0.042]	-0.0017[0.042]
0.5	0.4822[0.097]	0.4484[0.096]	0.4484[0.096]	0.4808[0.100]	0.4926[0.089]	0.4926[0.089]
1.0	0.9469[0.259]	0.8501[0.259]	0.8500[0.259]	0.9081[0.247]	0.9435[0.246]	0.9434[0.246]
1.0	0.9784[0.144]	1.0170[0.162]	1.0170[0.162]	0.9871[0.145]	0.9475[0.124]	0.9475[0.124]
	$n = 100, T = 3$					
5.0	4.9975[0.265]	4.9224[0.276]	4.9695[0.262]	4.6100[0.278]	5.0438[0.335]	5.1446[0.290]
1.0	1.0003[0.029]	0.9916[0.034]	0.9974[0.033]	0.9662[0.029]	1.0024[0.029]	1.0118[0.029]
1.0	1.0089[0.239]	0.9960[0.239]	1.0040[0.240]	0.9023[0.226]	0.9941[0.242]	1.0155[0.245]
<b>0.8</b>	0.8005[0.017]	0.8086[0.019]	0.8035[0.016]	0.8197[0.010]	0.7981[0.013]	0.7931[0.010]
0.5	0.4880[0.072]	0.2826[0.083]	0.2658[0.084]	0.4787[0.072]	0.4749[0.072]	0.4735[0.072]
1.0	0.9621[0.180]	0.1625[0.098]	0.0201[0.048]	0.8933[0.157]	0.9873[0.248]	0.9648[0.190]
1.0	0.9945[0.107]	1.2741[0.115]	1.2990[0.118]	1.0052[0.107]	0.9896[0.104]	0.9969[0.105]
5.0	4.9962[0.258]	4.8481[0.297]	4.8535[0.298]	4.7778[0.262]	5.0177[0.259]	5.0181[0.259]
1.0	1.0009[0.031]	0.9813[0.038]	0.9820[0.038]	0.9755[0.032]	1.0003[0.030]	1.0003[0.030]
1.0	1.0026[0.239]	0.9616[0.240]	0.9630[0.240]	0.9453[0.225]	0.9933[0.231]	0.9934[0.231]
<b>0.4</b>	0.4002[0.026]	0.4229[0.034]	0.4221[0.035]	0.4277[0.023]	0.3989[0.022]	0.3989[0.022]
0.5	0.4878[0.073]	0.3309[0.089]	0.3308[0.090]	0.4867[0.072]	0.4825[0.069]	0.4824[0.069]
1.0	0.9746[0.183]	0.4723[0.195]	0.4706[0.197]	0.9108[0.178]	0.9695[0.188]	0.9687[0.188]
1.0	0.9943[0.103]	1.1997[0.125]	1.2001[0.126]	1.0052[0.100]	0.9887[0.096]	0.9890[0.096]
5.0	4.9946[0.270]	5.0102[0.279]	5.0103[0.279]	4.9119[0.266]	5.0339[0.264]	5.0339[0.264]
1.0	0.9998[0.032]	0.9996[0.034]	0.9996[0.034]	0.9865[0.032]	1.0016[0.031]	1.0016[0.031]
1.0	1.0004[0.249]	0.9802[0.249]	0.9802[0.249]	0.9565[0.238]	0.9816[0.242]	0.9816[0.242]
<b>0.0</b>	0.0001[0.033]	-0.0008[0.036]	-0.0008[0.036]	0.0208[0.033]	-0.0032[0.031]	-0.0032[0.031]
0.5	0.4877[0.071]	0.4050[0.090]	0.4050[0.090]	0.4912[0.072]	0.5024[0.062]	0.5024[0.062]
1.0	0.9638[0.186]	0.8049[0.194]	0.8050[0.194]	0.9518[0.182]	0.9871[0.182]	0.9872[0.182]
1.0	0.9864[0.105]	1.0428[0.128]	1.0427[0.128]	0.9942[0.108]	0.9641[0.092]	0.9641[0.092]

**Note:**  $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \sigma_\mu, \sigma_v)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.01, .5, .5, 2, 1)$  (see Footnote 13).

**Table 2a.** Monte Carlo Mean and SD, and Bootstrap Standard Errors,  $m = 0$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
$n$	$\psi$	$T + 1 = 4$					$T + 1 = 8$				
		Normal Errors									
50	5.0	5.0155	0.3595	0.3257	0.3428	0.3759	5.0040	0.2736	0.2436	0.2695	0.3149
	1.0	1.0003	0.0422	0.0373	0.0403	0.0443	0.9999	0.0229	0.0203	0.0222	0.0246
	1.0	0.9949	0.3462	0.3321	0.3291	0.3288	0.9996	0.3017	0.2981	0.2978	0.2988
	0.5	0.4987	0.0332	0.0312	0.0321	0.0342	0.4995	0.0150	0.0140	0.0149	0.0162
	0.5	0.4775	0.1035	0.1037	0.1003	0.1104	0.4973	0.0608	0.0632	0.0588	0.063
	1.0	0.9998	0.3622	0.3885	0.3543	0.3692	0.9734	0.2657	0.2727	0.2543	0.2621
100	1.0	0.9775	0.1441	0.1416	0.1455	0.1686	0.9883	0.0822	0.0821	0.0837	0.0981
	5.0	5.0021	0.2634	0.2421	0.2571	0.2797	5.0014	0.1860	0.1591	0.1806	0.2145
	1.0	1.0000	0.0287	0.0270	0.0285	0.0305	1.0000	0.0155	0.0148	0.0160	0.0175
	1.0	0.9949	0.2412	0.2360	0.2350	0.2351	1.0109	0.2168	0.2141	0.2161	0.2190
	0.5	0.5000	0.0223	0.0211	0.0216	0.0226	0.4999	0.0105	0.0098	0.0105	0.0113
	0.5	0.4896	0.0726	0.0750	0.0715	0.0766	0.4976	0.0398	0.0466	0.0425	0.0444
100	1.0	1.0040	0.2540	0.2636	0.2495	0.2589	0.9866	0.1889	0.1871	0.1815	0.1885
	1.0	0.9899	0.1027	0.0964	0.1038	0.1227	0.9966	0.0602	0.0560	0.0596	0.0710
		Normal Mixture Errors									
50	5.0	5.0105	0.3450	0.3340	0.3389	0.3735	4.9986	0.2828	0.2555	0.2685	0.3100
	1.0	1.0005	0.0394	0.0368	0.0398	0.0441	1.0001	0.0208	0.0190	0.0205	0.0224
	1.0	0.9972	0.3300	0.3244	0.3215	0.3220	1.0029	0.3045	0.2977	0.2945	0.2928
	0.5	0.4997	0.0331	0.0308	0.0316	0.0345	0.4998	0.0159	0.0143	0.0149	0.0161
	0.5	0.4887	0.1011	0.0984	0.0985	0.1178	0.4928	0.0575	0.0584	0.0590	0.0719
	1.0	1.0376	0.6779	0.3104	0.3636	0.5621	1.0135	0.5932	0.1917	0.2625	0.4643
100	1.0	0.9813	0.2916	0.0897	0.1464	0.2867	0.9964	0.1770	0.0413	0.0844	0.1923
	5.0	5.0098	0.2541	0.2313	0.2420	0.2676	4.9899	0.1900	0.1671	0.1842	0.2175
	1.0	1.0002	0.0293	0.0272	0.0290	0.0316	0.9997	0.0154	0.0139	0.0151	0.0164
	1.0	0.9842	0.2397	0.2344	0.2310	0.2290	1.0070	0.2189	0.2115	0.2151	0.2197
	0.5	0.5004	0.0240	0.0208	0.0218	0.0236	0.5002	0.0106	0.0101	0.0106	0.0114
	0.5	0.4900	0.0696	0.0730	0.0713	0.0834	0.4972	0.0421	0.0440	0.0425	0.0502
100	1.0	1.0239	0.4462	0.1898	0.2532	0.4188	1.0078	0.3683	0.1162	0.1850	0.3578
	1.0	0.9927	0.2081	0.0569	0.1042	0.2177	0.9901	0.1289	0.0265	0.0592	0.1416
		Chi-Square Errors, df=5									
50	5.0	4.9959	0.3544	0.3414	0.3420	0.3756	5.0178	0.3216	0.3135	0.3190	0.3535
	1.0	0.9994	0.0408	0.0373	0.0403	0.0443	1.0006	0.0236	0.0220	0.0231	0.0246
	1.0	0.9942	0.3366	0.3318	0.3287	0.3285	0.9943	0.3363	0.3330	0.3286	0.3258
	0.5	0.5017	0.0334	0.0307	0.0320	0.0350	0.4982	0.0154	0.0148	0.0153	0.0163
	0.5	0.4758	0.1012	0.1026	0.1005	0.1133	0.4959	0.0582	0.0615	0.0588	0.0651
	1.0	1.0195	0.4533	0.3659	0.3601	0.4293	0.9649	0.3417	0.2488	0.2527	0.3186
100	1.0	0.9806	0.1876	0.1208	0.1460	0.2072	0.9895	0.1166	0.0631	0.0838	0.1273
	5.0	4.9997	0.2478	0.2430	0.2455	0.2691	4.9919	0.1903	0.1788	0.1885	0.2209
	1.0	0.9997	0.0286	0.0262	0.0282	0.0308	0.9993	0.0156	0.0143	0.0155	0.0169
	1.0	0.9981	0.2343	0.2359	0.2352	0.2357	1.0062	0.2157	0.2116	0.2126	0.2143
	0.5	0.5002	0.0216	0.0204	0.0214	0.0229	0.5002	0.0110	0.0104	0.0110	0.0118
	0.5	0.4889	0.0673	0.0744	0.0716	0.0787	0.4974	0.0426	0.0455	0.0425	0.0458
100	1.0	1.0103	0.3043	0.2381	0.2501	0.3066	0.9824	0.2466	0.1653	0.1810	0.2397
	1.0	0.9917	0.1391	0.0799	0.1040	0.1536	0.9946	0.0838	0.0421	0.0595	0.0934

**Note:**  $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \phi_\mu, \sigma_v^2)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.01, .5, .5, 2, 1)$  (see Footnote 13).

**Table 2b.** Monte Carlo Mean and SD, and Bootstrap Standard Errors,  $m = 6$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb	
$n$	$\psi$	$T + 1 = 4$					$T + 1 = 8$					
		Normal Errors										
50	5.0	5.0006	0.3692	0.3683	0.3677	0.3947	5.0104	0.2857	0.2931	0.2770	0.3033	
	1.0	0.9989	0.0371	0.0364	0.0378	0.0408	1.0014	0.0247	0.0253	0.0251	0.0264	
	1.0	0.9489	0.3510	0.3637	0.3626	0.3732	0.9917	0.3106	0.3047	0.2986	0.3001	
	0.5	0.5014	0.0275	0.0289	0.0277	0.0281	0.4990	0.0151	0.0206	0.0153	0.0121	
	0.5	0.4972	0.0907	0.0953	0.0906	0.1004	0.4832	0.0601	0.0616	0.0583	0.0637	
	1.0	0.9905	0.3505	0.3737	0.3424	0.3635	0.9678	0.2583	0.2832	0.2534	0.2584	
100	1.0	0.9805	0.1439	0.1381	0.1425	0.1687	0.9900	0.0872	0.0828	0.0835	0.0989	
	5.0	5.0276	0.2902	0.2687	0.2739	0.2910	5.0036	0.2046	0.2037	0.1966	0.2126	
	1.0	1.0017	0.0297	0.0285	0.0296	0.0314	1.0005	0.0163	0.0163	0.0163	0.0170	
	1.0	1.0203	0.2406	0.2402	0.2351	0.2331	0.9996	0.2197	0.2158	0.2128	0.2130	
	0.5	0.4973	0.0212	0.0209	0.0203	0.0203	0.4997	0.0109	0.0140	0.0112	0.0094	
	0.5	0.4898	0.0681	0.0714	0.0676	0.0718	0.4966	0.0412	0.0451	0.0414	0.0436	
	1.0	1.0103	0.2643	0.2666	0.2537	0.2649	0.9836	0.1796	0.1915	0.1816	0.1877	
	1.0	0.9879	0.1020	0.0946	0.1015	0.1203	0.9948	0.0579	0.0559	0.0594	0.0710	
			Normal Mixture Errors									
	50	5.0	5.0188	0.3582	0.3763	0.3684	0.4236	5.0123	0.2804	0.3036	0.2777	0.3024
		1.0	1.0003	0.0383	0.0364	0.0378	0.0434	1.0013	0.0259	0.0252	0.0250	0.0263
		1.0	0.9170	0.3839	0.3591	0.3579	0.3835	0.9963	0.2960	0.3064	0.2996	0.3004
0.5		0.5010	0.0282	0.0287	0.0281	0.0324	0.4991	0.0155	0.0205	0.0152	0.0121	
0.5		0.4941	0.0903	0.0922	0.0907	0.1096	0.4856	0.0567	0.0571	0.0581	0.0732	
1.0		1.0256	0.6788	0.3003	0.3543	0.5729	1.0370	0.5664	0.2124	0.2691	0.4816	
100	1.0	0.9938	0.2765	0.0843	0.1461	0.3087	0.9911	0.1791	0.0416	0.0836	0.1925	
	5.0	5.0199	0.2863	0.2722	0.2734	0.2941	4.9971	0.1975	0.2075	0.1960	0.2116	
	1.0	1.0014	0.0295	0.0283	0.0294	0.0316	1.0003	0.0161	0.0163	0.0162	0.0170	
	1.0	1.0066	0.2531	0.2387	0.2336	0.2319	1.0082	0.2109	0.2147	0.2116	0.2116	
	0.5	0.4983	0.0206	0.0207	0.0202	0.0208	0.4997	0.0113	0.0139	0.0111	0.0094	
	0.5	0.4905	0.0672	0.0695	0.0675	0.0795	0.4969	0.0397	0.0428	0.0415	0.0496	
	1.0	1.0475	0.4597	0.2037	0.2626	0.4341	1.0091	0.4092	0.1281	0.1855	0.3568	
	1.0	0.9837	0.2014	0.0537	0.1014	0.2178	0.9943	0.1302	0.0270	0.0593	0.1416	
			Chi-Square Errors, df=5									
	50	5.0	5.0165	0.3750	0.3859	0.3697	0.3991	5.0351	0.2870	0.3065	0.2770	0.3015
		1.0	0.9984	0.0383	0.0365	0.0378	0.0411	1.0013	0.0255	0.0251	0.0250	0.0263
		1.0	0.9227	0.3595	0.3633	0.3621	0.3754	0.9583	0.3014	0.3049	0.2985	0.2996
0.5		0.5008	0.0277	0.0289	0.0278	0.0288	0.4992	0.0148	0.0205	0.0152	0.0120	
0.5		0.5031	0.0877	0.0938	0.0900	0.1028	0.4849	0.0584	0.0601	0.0582	0.0662	
1.0		0.9992	0.4431	0.3510	0.3446	0.4179	0.9925	0.3520	0.2700	0.2590	0.3251	
100	1.0	0.9906	0.1940	0.1202	0.1441	0.2107	0.9833	0.1181	0.0638	0.0829	0.1281	
	5.0	5.0307	0.2801	0.2807	0.2744	0.2908	5.0081	0.1999	0.2133	0.1967	0.2119	
	1.0	1.0016	0.0296	0.0285	0.0296	0.0315	1.0004	0.0169	0.0163	0.0163	0.0170	
	1.0	1.0172	0.2419	0.2405	0.2358	0.2343	0.9989	0.2137	0.2157	0.2128	0.2130	
	0.5	0.4969	0.0203	0.0208	0.0203	0.0205	0.4996	0.0112	0.0140	0.0112	0.0094	
	0.5	0.4888	0.0689	0.0709	0.0677	0.0741	0.4960	0.0426	0.0443	0.0415	0.0452	
	1.0	1.0304	0.3157	0.2479	0.2584	0.3169	0.9949	0.2548	0.1757	0.1833	0.2396	
	1.0	0.9867	0.1323	0.0791	0.1015	0.1512	0.9932	0.0810	0.0430	0.0593	0.0931	

**Note:**  $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \phi_\mu, \sigma_v^2)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.01, .5, .5, 2, 1)$  (see Footnote 13).

**Table 2c.** Monte Carlo Mean and SD, and Bootstrap Standard Errors,  $m = 200$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
$n$	$\psi$	$T + 1 = 4$					$T + 1 = 8$				
Normal Errors											
50	5.0	5.0283	0.3738	0.3745	0.3731	0.3958	5.0117	0.2852	0.2966	0.2834	0.3117
	1.0	1.0012	0.0392	0.0387	0.0397	0.0423	1.0003	0.0250	0.0248	0.0237	0.0243
	1.0	0.9720	0.3411	0.3339	0.3321	0.3369	1.0028	0.3041	0.3033	0.3046	0.3130
	0.5	0.4970	0.0275	0.0280	0.0265	0.0263	0.4993	0.0157	0.0217	0.0162	0.0129
	0.5	0.4778	0.0907	0.0981	0.0934	0.1017	0.4922	0.0599	0.0611	0.0575	0.0627
	1.0	1.0255	0.3967	0.3912	0.3602	0.3833	0.9842	0.2643	0.2863	0.2576	0.2646
100	1.0	0.9742	0.1484	0.1380	0.1424	0.1685	0.9898	0.0817	0.0825	0.0835	0.0991
	5.0	5.0121	0.2733	0.2740	0.2727	0.2849	5.0113	0.2131	0.2116	0.2059	0.2254
	1.0	1.0001	0.0305	0.0287	0.0298	0.0316	1.0006	0.0177	0.0176	0.0176	0.0185
	1.0	1.0020	0.2423	0.2421	0.2418	0.2448	0.9853	0.2247	0.2155	0.2137	0.2149
	0.5	0.4988	0.0213	0.0218	0.0205	0.0199	0.5000	0.0120	0.0150	0.0117	0.0095
	0.5	0.4963	0.0663	0.0707	0.0667	0.0707	0.4989	0.0408	0.0452	0.0417	0.0438
100	1.0	1.0026	0.2702	0.2679	0.2535	0.2638	0.9747	0.1845	0.1934	0.1813	0.1854
	1.0	0.9865	0.1024	0.0938	0.1015	0.1212	0.9985	0.0605	0.0564	0.0597	0.0711
Normal Mixture Errors											
50	5.0	5.0122	0.3683	0.3803	0.3677	0.4082	5.0039	0.2902	0.3019	0.2799	0.3079
	1.0	0.9986	0.0412	0.0385	0.0395	0.0437	1.0001	0.0238	0.0247	0.0235	0.0241
	1.0	0.9767	0.3368	0.3274	0.3248	0.3312	1.0178	0.3164	0.2979	0.2987	0.3066
	0.5	0.4993	0.0263	0.0275	0.0263	0.0285	0.4995	0.0161	0.0214	0.0160	0.0130
	0.5	0.4707	0.0960	0.0948	0.0938	0.1130	0.4945	0.0585	0.0566	0.0573	0.0711
	1.0	1.0508	0.7028	0.3138	0.3660	0.5834	1.0052	0.5478	0.2101	0.2621	0.4621
100	1.0	0.9808	0.2897	0.0855	0.1438	0.2965	0.9855	0.1855	0.0417	0.0832	0.1900
	5.0	4.9976	0.2751	0.2757	0.2705	0.2861	5.0239	0.2076	0.2165	0.2058	0.2248
	1.0	1.0018	0.0304	0.0286	0.0296	0.0316	1.0000	0.0178	0.0176	0.0176	0.0185
	1.0	0.9985	0.2392	0.2392	0.2390	0.2422	0.9823	0.2159	0.2151	0.2127	0.2136
	0.5	0.5004	0.0208	0.0216	0.0204	0.0203	0.4992	0.0118	0.0150	0.0117	0.0096
	0.5	0.4933	0.0670	0.0690	0.0669	0.0781	0.5003	0.0408	0.0429	0.0416	0.0495
100	1.0	1.0146	0.4514	0.2034	0.2555	0.4149	0.9902	0.3572	0.1302	0.1840	0.3490
	1.0	0.9863	0.1955	0.0547	0.1017	0.2159	1.0014	0.1294	0.0272	0.0599	0.1440
Chi-Square Errors, df=5											
50	5.0	5.0403	0.3978	0.3932	0.3732	0.3927	5.0213	0.2890	0.3071	0.2811	0.3075
	1.0	0.9996	0.0405	0.0386	0.0396	0.0423	1.0007	0.0238	0.0247	0.0236	0.0242
	1.0	0.9744	0.3420	0.3345	0.3317	0.3358	1.0090	0.3283	0.2997	0.3014	0.3098
	0.5	0.4972	0.0264	0.0280	0.0264	0.0263	0.4983	0.0162	0.0216	0.0161	0.0128
	0.5	0.4766	0.0912	0.0976	0.0935	0.1041	0.4931	0.0586	0.0595	0.0574	0.0648
	1.0	1.0448	0.4633	0.3701	0.3627	0.4375	0.9824	0.3657	0.2678	0.2568	0.3208
100	1.0	0.9703	0.1867	0.1194	0.1414	0.2023	0.9853	0.1162	0.0651	0.0831	0.1257
	5.0	4.9983	0.2807	0.2860	0.2728	0.2836	5.0051	0.2098	0.2210	0.2059	0.2244
	1.0	1.0023	0.0299	0.0287	0.0298	0.0316	1.0001	0.0178	0.0176	0.0176	0.0185
	1.0	1.0055	0.2416	0.2425	0.2418	0.2443	0.9941	0.2150	0.2161	0.2139	0.2147
	0.5	0.4996	0.0212	0.0218	0.0205	0.0200	0.4998	0.0119	0.0150	0.0117	0.0095
	0.5	0.4995	0.0647	0.0700	0.0666	0.0725	0.4989	0.0400	0.0444	0.0417	0.0454
100	1.0	1.0081	0.3351	0.2480	0.2542	0.3083	0.9862	0.2441	0.1769	0.1835	0.2393
	1.0	0.9921	0.1389	0.0798	0.1021	0.1514	0.9965	0.0805	0.0429	0.0596	0.0942

**Note:**  $\psi = (\gamma_0, \beta, \gamma_1, \rho, \lambda, \phi_\mu, \sigma_v^2)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.01, .5, .5, 2, 1)$  (see Footnote 13).



**Table 3a.** Monte Carlo Mean[RMSE] for the QMLEs, Fixed Effects Model, Normal Errors

$\psi$	true $m = 0$			true $m = 6$		
	$m = 0$	$m = 6$	$m = 200$	$m = 0$	$m = 6$	$m = 200$
$n = 50, T = 3$						
1.0	0.9957[.090]	0.9702[.088]	0.9589[.087]	1.0006[.127]	0.9983[.126]	0.9891[.125]
<b>-0.9</b>	-0.8966[.045]	-0.8390[.038]	-0.8139[.029]	-0.8976[.037]	-0.8934[.034]	-0.8744[.026]
0.5	0.4764[.105]	0.4471[.100]	0.4584[.100]	0.4912[.104]	0.4889[.088]	0.4837[.088]
1.0	0.9775[.141]	0.8568[.113]	0.8747[.116]	0.9934[.132]	0.9632[.131]	0.9521[.131]
1.0	0.9989[.089]	0.9969[.089]	0.9969[.089]	0.9934[.135]	0.9926[.133]	0.9926[.133]
<b>-0.5</b>	-0.4996[.048]	-0.4926[.048]	-0.4925[.048]	-0.4943[.074]	-0.4924[.068]	-0.4923[.068]
0.5	0.4852[.102]	0.4092[.117]	0.4091[.117]	0.5149[.114]	0.4893[.095]	0.4893[.095]
1.0	0.9662[.142]	0.9493[.142]	0.9493[.142]	0.9734[.153]	0.9410[.136]	0.9410[.136]
1.0	0.9991[.090]	0.9990[.090]	0.9990[.090]	0.9904[.139]	1.0012[.136]	1.0012[.136]
<b>0.0</b>	0.0004[.055]	-0.0004[.055]	-0.0004[.055]	0.0280[.103]	-0.0059[.087]	-0.0059[.087]
0.5	0.4925[.100]	0.4780[.097]	0.4780[.097]	0.5281[.101]	0.4903[.089]	0.4903[.089]
1.0	0.9673[.149]	0.9619[.147]	0.9619[.147]	1.0134[.176]	0.9340[.130]	0.9340[.130]
1.0	0.9988[.095]	0.9989[.095]	0.9988[.095]	1.0031[.135]	1.0049[.134]	1.0050[.134]
<b>0.5</b>	0.4976[.040]	0.4977[.040]	0.4977[.040]	0.5155[.096]	0.4983[.089]	0.4982[.089]
0.5	0.4772[.108]	0.4675[.107]	0.4675[.107]	0.5081[.102]	0.4826[.098]	0.4826[.098]
1.0	0.9610[.144]	0.9586[.144]	0.9586[.144]	0.9973[.174]	0.9703[.156]	0.9702[.156]
1.0	1.0035[.089]	1.0037[.089]	1.0037[.089]	0.9977[.133]	0.9976[.133]	0.9976[.133]
<b>0.9</b>	0.8991[.025]	0.8993[.025]	0.8993[.025]	0.9004[.044]	0.9002[.044]	0.9002[.044]
0.5	0.4704[.112]	0.4695[.112]	0.4692[.112]	0.4862[.104]	0.4859[.103]	0.4858[.103]
1.0	0.9682[.149]	0.9682[.149]	0.9681[.149]	0.9803[.151]	0.9803[.151]	0.9803[.151]
$n = 100, T = 3$						
1.0	1.0025[.074]	0.9882[.074]	0.9750[.073]	0.9986[.071]	0.9985[.071]	0.9935[.071]
<b>-0.9</b>	-0.8996[.026]	-0.8753[.023]	-0.8528[.017]	-0.8996[.026]	-0.8994[.024]	-0.8858[.019]
0.5	0.4937[.077]	0.3917[.075]	0.4014[.073]	0.5001[.076]	0.4876[.068]	0.4753[.068]
1.0	0.9848[.104]	0.9411[.089]	0.9410[.091]	1.0177[.093]	0.9847[.102]	0.9765[.098]
1.0	0.9972[.075]	0.9951[.075]	0.9950[.075]	0.9994[.071]	1.0007[.070]	1.0006[.070]
<b>-0.5</b>	-0.5026[.038]	-0.4977[.037]	-0.4976[.037]	-0.4951[.050]	-0.4983[.047]	-0.4983[.047]
0.5	0.4892[.076]	0.4289[.078]	0.4289[.078]	0.5302[.081]	0.4977[.065]	0.4977[.065]
1.0	0.9790[.107]	0.9696[.106]	0.9696[.106]	0.9984[.107]	0.9792[.098]	0.9792[.098]
1.0	0.9992[.076]	0.9997[.075]	0.9997[.075]	0.9941[.072]	1.0022[.071]	1.0022[.071]
<b>0.0</b>	0.0022[.041]	0.0011[.041]	0.0011[.041]	0.0223[.064]	-0.0072[.056]	-0.0072[.056]
0.5	0.4989[.073]	0.4848[.068]	0.4848[.068]	0.5472[.075]	0.4977[.063]	0.4977[.063]
1.0	0.9944[.106]	0.9916[.105]	0.9916[.105]	1.0225[.119]	0.9584[.091]	0.9584[.091]
1.0	0.9989[.075]	0.9989[.075]	0.9989[.075]	0.9997[.069]	1.0001[.069]	1.0001[.069]
<b>0.5</b>	0.5014[.031]	0.5012[.030]	0.5012[.030]	0.5188[.062]	0.5036[.057]	0.5036[.057]
0.5	0.5001[.077]	0.4969[.076]	0.4969[.076]	0.5193[.070]	0.4957[.067]	0.4957[.067]
1.0	0.9829[.106]	0.9827[.106]	0.9827[.106]	1.0224[.122]	1.0056[.113]	1.0056[.113]
1.0	0.9952[.071]	0.9952[.071]	0.9952[.071]	0.9990[.068]	0.9991[.068]	0.9991[.068]
<b>0.9</b>	0.9003[.021]	0.9001[.021]	0.9002[.021]	0.9018[.028]	0.9020[.028]	0.9020[.028]
0.5	0.4952[.077]	0.4954[.077]	0.4954[.077]	0.4864[.076]	0.4857[.075]	0.4855[.075]
1.0	0.9844[.108]	0.9843[.108]	0.9843[.108]	0.9834[.104]	0.9836[.104]	0.9835[.104]

**Note:**  $\psi = (\beta, \rho, \lambda, \sigma_v)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.01, .5, .5, 1, .5)$  (see Footnote 13).

**Table 3b.** Monte Carlo Mean[RMSE] for the QMLEs, Fixed Effects Model, Normal Mixture

$\psi$	true $m = 0$			true $m = 6$		
	$m = 0$	$m = 6$	$m = 200$	$m = 0$	$m = 6$	$m = 200$
$n = 50, T = 3$						
1.0	1.0021[.092]	0.9906[.091]	0.9826[.090]	0.9981[.126]	0.9980[.125]	0.9954[.125]
<b>-0.9</b>	-0.8987[.041]	-0.8648[.040]	-0.8416[.033]	-0.8956[.038]	-0.8924[.038]	-0.8770[.033]
0.5	0.4862[.103]	0.4035[.098]	0.4113[.097]	0.4829[.105]	0.4850[.092]	0.4770[.091]
1.0	0.9822[.300]	0.9147[.252]	0.9238[.262]	1.0121[.264]	0.9540[.268]	0.9473[.271]
1.0	1.0026[.091]	1.0013[.091]	1.0013[.091]	0.9923[.128]	0.9905[.127]	0.9905[.127]
<b>-0.5</b>	-0.5009[.050]	-0.4969[.049]	-0.4969[.049]	-0.4926[.079]	-0.4881[.072]	-0.4880[.072]
0.5	0.4894[.103]	0.4415[.103]	0.4415[.103]	0.5164[.103]	0.4934[.089]	0.4934[.089]
1.0	0.9802[.285]	0.9687[.278]	0.9687[.278]	0.9807[.291]	0.9301[.247]	0.9301[.247]
1.0	0.9986[.089]	0.9986[.089]	0.9986[.089]	0.9936[.139]	1.0045[.134]	1.0045[.134]
<b>0.0</b>	0.0017[.062]	0.0005[.062]	0.0005[.062]	0.0254[.106]	-0.0110[.091]	-0.0110[.091]
0.5	0.4917[.102]	0.4733[.098]	0.4733[.098]	0.5371[.099]	0.5045[.088]	0.5045[.088]
1.0	0.9761[.305]	0.9731[.302]	0.9731[.302]	1.0100[.309]	0.9057[.235]	0.9057[.235]
1.0	1.0004[.090]	1.0004[.090]	1.0004[.090]	1.0033[.129]	1.0051[.128]	1.0051[.128]
<b>0.5</b>	0.5001[.041]	0.5000[.041]	0.5000[.041]	0.5068[.100]	0.4911[.094]	0.4911[.094]
0.5	0.4826[.105]	0.4761[.104]	0.4761[.104]	0.5054[.097]	0.4809[.094]	0.4808[.094]
1.0	0.9865[.303]	0.9844[.301]	0.9844[.301]	0.9824[.313]	0.9551[.287]	0.9550[.286]
1.0	0.9968[.094]	0.9970[.094]	0.9970[.094]	0.9971[.128]	0.9970[.128]	0.9970[.128]
<b>0.9</b>	0.8991[.026]	0.8993[.026]	0.8993[.026]	0.9006[.049]	0.9004[.049]	0.9004[.049]
0.5	0.4797[.107]	0.4789[.107]	0.4786[.107]	0.4884[.106]	0.4881[.105]	0.4880[.105]
1.0	0.9760[.279]	0.9760[.279]	0.9759[.279]	0.9649[.285]	0.9648[.284]	0.9649[.284]
$n = 100, T = 3$						
1.0	0.9986[.076]	0.9712[.075]	0.9564[.074]	1.0022[.072]	1.0028[.072]	0.9979[.072]
<b>-0.9</b>	-0.9005[.030]	-0.8549[.029]	-0.8303[.023]	-0.8964[.026]	-0.8972[.025]	-0.8853[.021]
0.5	0.4909[.078]	0.4299[.071]	0.4398[.072]	0.4938[.074]	0.4864[.068]	0.4744[.068]
1.0	0.9833[.205]	0.8850[.164]	0.8978[.173]	1.0367[.177]	0.9845[.200]	0.9779[.198]
1.0	0.9976[.074]	0.9964[.074]	0.9964[.074]	0.9971[.073]	0.9971[.072]	0.9971[.072]
<b>-0.5</b>	-0.4987[.039]	-0.4963[.039]	-0.4963[.039]	-0.4922[.055]	-0.4926[.052]	-0.4925[.052]
0.5	0.5002[.080]	0.4672[.074]	0.4672[.074]	0.5262[.076]	0.4967[.062]	0.4967[.062]
1.0	0.9862[.204]	0.9742[.200]	0.9742[.200]	0.9994[.219]	0.9641[.188]	0.9641[.188]
1.0	1.0016[.077]	1.0017[.077]	1.0017[.077]	0.9930[.073]	1.0011[.072]	1.0011[.072]
<b>0.0</b>	-0.0014[.038]	-0.0015[.038]	-0.0015[.038]	0.0229[.067]	-0.0072[.059]	-0.0072[.059]
0.5	0.4921[.073]	0.4694[.071]	0.4694[.071]	0.5428[.074]	0.4998[.064]	0.4998[.064]
1.0	0.9892[.208]	0.9864[.207]	0.9864[.207]	1.0143[.224]	0.9344[.175]	0.9344[.175]
1.0	1.0003[.074]	1.0005[.074]	1.0005[.074]	1.0005[.070]	1.0010[.069]	1.0010[.069]
<b>0.5</b>	0.5012[.033]	0.5005[.032]	0.5005[.032]	0.5201[.067]	0.5050[.062]	0.5050[.062]
0.5	0.5131[.076]	0.5162[.073]	0.5162[.073]	0.5174[.067]	0.4941[.063]	0.4941[.063]
1.0	0.9912[.218]	0.9912[.218]	0.9912[.218]	1.0245[.222]	1.0047[.204]	1.0047[.204]
1.0	1.0019[.073]	1.0019[.073]	1.0019[.073]	0.9976[.076]	0.9977[.076]	0.9977[.076]
<b>0.9</b>	0.9005[.021]	0.9003[.021]	0.9003[.021]	0.9011[.028]	0.9013[.028]	0.9013[.028]
0.5	0.4976[.079]	0.4979[.079]	0.4980[.079]	0.4853[.076]	0.4846[.075]	0.4843[.075]
1.0	0.9816[.205]	0.9814[.204]	0.9815[.204]	0.9801[.202]	0.9803[.202]	0.9802[.202]

**Note:**  $\psi = (\beta, \rho, \lambda, \sigma_v)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.01, .5, .5, 1, .5)$  (see Footnote 13).

**Table 4a.** Monte Carlo Mean and SD, and Bootstrap Standard Errors,  $m = 0$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
$n$	$\psi$	$T = 3$					$T = 7$				
		Normal Errors									
50	1.0	0.9986	0.0971	0.1001	0.0981	0.0982	1.0003	0.0559	0.0545	0.0532	0.0549
	0.5	0.4988	0.0348	0.0380	0.0326	0.0437	0.4995	0.0241	0.0259	0.0241	0.0363
	0.5	0.4888	0.1055	0.1016	0.1044	0.1127	0.4917	0.0612	0.0571	0.0597	0.0639
	1.0	0.9650	0.1489	0.1713	0.1411	0.1339	0.9861	0.0806	0.0990	0.0841	0.0794
100	1.0	1.0024	0.0720	0.0744	0.0737	0.0790	1.0005	0.0340	0.0343	0.0337	0.0342
	0.5	0.5012	0.0266	0.0288	0.0273	0.0417	0.5005	0.0167	0.0173	0.0170	0.0266
	0.5	0.4922	0.0759	0.0742	0.0749	0.0782	0.4986	0.0408	0.0419	0.0428	0.0443
	1.0	0.9889	0.1044	0.1219	0.1022	0.0980	0.9948	0.0592	0.0673	0.0600	0.0576
		Normal Mixture Errors									
50	1.0	0.9979	0.0967	0.0996	0.0971	0.0973	1.0016	0.0530	0.0550	0.0533	0.0563
	0.5	0.4976	0.0338	0.0385	0.0320	0.0461	0.4994	0.0252	0.0278	0.0249	0.0408
	0.5	0.4847	0.1017	0.1001	0.1046	0.1153	0.4953	0.0585	0.0542	0.0595	0.0671
	1.0	0.9586	0.2841	0.1207	0.1401	0.2372	0.9881	0.1855	0.0637	0.0844	0.1610
100	1.0	1.0027	0.0733	0.0742	0.0733	0.0791	0.9971	0.0328	0.0342	0.0336	0.0341
	0.5	0.5000	0.0269	0.0287	0.0262	0.0431	0.4994	0.0168	0.0173	0.0169	0.0275
	0.5	0.4933	0.0718	0.0731	0.0748	0.0794	0.4995	0.0435	0.0406	0.0428	0.0457
	1.0	0.9860	0.2121	0.0833	0.1019	0.1860	0.9894	0.1291	0.0408	0.0596	0.1198
		Chi-Square, df=3									
50	1.0	0.9942	0.1022	0.1001	0.0983	0.0995	1.0034	0.0544	0.0549	0.0534	0.0557
	0.5	0.4999	0.0361	0.0376	0.0333	0.0471	0.4991	0.0251	0.0265	0.0242	0.0369
	0.5	0.4785	0.1046	0.1015	0.1060	0.1171	0.4966	0.0588	0.0554	0.0595	0.0654
	1.0	0.9646	0.2141	0.1377	0.1409	0.1860	0.9908	0.1365	0.0741	0.0845	0.1218
100	1.0	1.0012	0.0734	0.0744	0.0737	0.0792	1.0010	0.0328	0.0344	0.0338	0.0345
	0.5	0.4999	0.0312	0.0290	0.0284	0.0487	0.5003	0.0175	0.0168	0.0169	0.0263
	0.5	0.4935	0.0771	0.0735	0.0755	0.0804	0.4976	0.0441	0.0414	0.0428	0.0449
	1.0	0.9918	0.1604	0.0971	0.1024	0.1425	0.9962	0.0971	0.0486	0.0600	0.0897

**Note:**  $\psi = (\beta, \rho, \lambda, \sigma_v^2)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.1, .5, .5, 5, 1)$  (see Footnote 13).

**Table 4b.** Monte Carlo Mean and SD, and Bootstrap Standard Errors,  $m = 6$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
$n$	$\psi$	$T = 3$					$T = 7$				
Normal Errors											
50	1.0	1.0000	0.0182	0.0189	0.0184	0.0183	1.0004	0.0095	0.0098	0.0096	0.0117
	0.5	0.5010	0.0198	0.0188	0.0190	0.0229	0.5001	0.0070	0.0073	0.0070	0.0089
	0.5	0.5000	0.1037	0.0999	0.1016	0.1058	0.4956	0.0603	0.0565	0.0594	0.0633
	1.0	0.9744	0.1450	0.1602	0.1427	0.1358	0.9914	0.0814	0.0907	0.0836	0.0809
100	1.0	0.9998	0.0150	0.0151	0.0149	0.0148	0.9999	0.0064	0.0068	0.0066	0.0075
	0.5	0.4992	0.0108	0.0117	0.0112	0.0121	0.5000	0.0052	0.0051	0.0051	0.0060
	0.5	0.4954	0.0701	0.0735	0.0728	0.0730	0.4991	0.0433	0.0418	0.0425	0.0437
	1.0	0.9805	0.1040	0.1082	0.1013	0.0990	0.9916	0.0638	0.0619	0.0591	0.0581
Normal Mixture Errors											
50	1.0	1.0004	0.0186	0.0187	0.0180	0.0179	0.9996	0.0093	0.0098	0.0095	0.0117
	0.5	0.4999	0.0196	0.0185	0.0187	0.0235	0.4999	0.0067	0.0073	0.0069	0.0089
	0.5	0.4993	0.1029	0.0978	0.1019	0.1090	0.4977	0.0572	0.0537	0.0592	0.0662
	1.0	0.9558	0.2840	0.0986	0.1400	0.2405	0.9857	0.1872	0.0471	0.0832	0.1677
100	1.0	0.9993	0.0156	0.0151	0.0149	0.0149	1.0000	0.0067	0.0067	0.0066	0.0074
	0.5	0.4997	0.0119	0.0117	0.0112	0.0128	0.4998	0.0049	0.0051	0.0051	0.0060
	0.5	0.4948	0.0719	0.0726	0.0729	0.0741	0.4976	0.0438	0.0407	0.0426	0.0451
	1.0	0.9906	0.2015	0.0647	0.1024	0.1908	0.9897	0.1301	0.0317	0.0590	0.1243
Chi-Square, df=3											
50	1.0	0.9991	0.0187	0.0189	0.0183	0.0182	1.0001	0.0100	0.0099	0.0096	0.0118
	0.5	0.4994	0.0195	0.0186	0.0189	0.0232	0.4998	0.0072	0.0074	0.0070	0.0089
	0.5	0.4958	0.0998	0.0997	0.1022	0.1071	0.4981	0.0569	0.0552	0.0593	0.0646
	1.0	0.9691	0.2161	0.1221	0.1418	0.1884	0.9995	0.1353	0.0615	0.0844	0.1269
100	1.0	1.0007	0.0146	0.0151	0.0149	0.0148	1.0000	0.0067	0.0068	0.0066	0.0075
	0.5	0.4999	0.0115	0.0117	0.0112	0.0124	0.4998	0.0049	0.0051	0.0051	0.0060
	0.5	0.4919	0.0704	0.0734	0.0732	0.0740	0.4977	0.0425	0.0414	0.0426	0.0443
	1.0	0.9811	0.1476	0.0803	0.1014	0.1418	0.9959	0.0955	0.0415	0.0594	0.0912

**Note:**  $\psi = (\beta, \rho, \lambda, \sigma_v^2)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.1, .5, .5, 5, 1)$  (see Footnote 13)

**Table 4c.** Monte Carlo Mean and SD, and Bootstrap Standard Errors,  $m = 200$

		Mean	SD	seSCb	seHS	seHSb	Mean	SD	seSCb	seHS	seHSb
$n$	$\psi$	$T = 3$					$T = 7$				
Normal Errors											
50	1.0	1.0004	0.0210	0.0213	0.0208	0.0210	1.0000	0.0097	0.0096	0.0093	0.0100
	0.5	0.4999	0.0197	0.0199	0.0197	0.0231	0.5000	0.0070	0.0072	0.0069	0.0081
	0.5	0.4866	0.0974	0.1011	0.1009	0.1027	0.4991	0.0626	0.0562	0.0588	0.0622
	1.0	0.9624	0.1422	0.1573	0.1406	0.1349	0.9909	0.0881	0.0914	0.0837	0.0800
100	1.0	1.0001	0.0139	0.0140	0.0138	0.0154	0.9990	0.0337	0.0339	0.0333	0.0358
	0.5	0.5001	0.0117	0.0117	0.0116	0.0144	0.4986	0.0201	0.0195	0.0206	0.0370
	0.5	0.4977	0.0736	0.0726	0.0745	0.0775	0.4991	0.0409	0.0397	0.0409	0.0430
	1.0	0.9886	0.1064	0.1091	0.1019	0.0993	0.9938	0.0585	0.0673	0.0601	0.0582
Normal Mixture Errors											
50	1.0	1.0005	0.0208	0.0213	0.0207	0.0210	0.9996	0.0092	0.0095	0.0092	0.0100
	0.5	0.4999	0.0204	0.0200	0.0196	0.0244	0.4997	0.0069	0.0072	0.0069	0.0082
	0.5	0.4796	0.1010	0.0994	0.1017	0.1064	0.5014	0.0566	0.0534	0.0586	0.0653
	1.0	0.9685	0.2847	0.1000	0.1414	0.2444	0.9937	0.1837	0.0474	0.0840	0.1685
100	1.0	1.0001	0.0138	0.0139	0.0137	0.0153	0.9994	0.0328	0.0339	0.0333	0.0360
	0.5	0.5000	0.0117	0.0117	0.0115	0.0148	0.5006	0.0209	0.0194	0.0205	0.0403
	0.5	0.4988	0.0743	0.0714	0.0743	0.0785	0.4967	0.0408	0.0387	0.0410	0.0445
	1.0	0.9835	0.2065	0.0642	0.1013	0.1879	0.9933	0.1339	0.0430	0.0600	0.1200
Chi-Square, df=3											
50	1.0	1.0002	0.0214	0.0213	0.0208	0.0211	1.0000	0.0094	0.0096	0.0093	0.0099
	0.5	0.4995	0.0203	0.0199	0.0197	0.0238	0.5001	0.0069	0.0072	0.0070	0.0081
	0.5	0.4835	0.1009	0.1003	0.1014	0.1048	0.4990	0.0549	0.0550	0.0587	0.0634
	1.0	0.9662	0.2116	0.1220	0.1411	0.1879	0.9944	0.1367	0.0614	0.0840	0.1255
100	1.0	1.0002	0.0144	0.0139	0.0137	0.0153	1.0009	0.0335	0.0338	0.0333	0.0359
	0.5	0.5005	0.0113	0.0117	0.0115	0.0145	0.4999	0.0207	0.0193	0.0205	0.0375
	0.5	0.4987	0.0732	0.0721	0.0744	0.0780	0.5004	0.0407	0.0392	0.0408	0.0435
	1.0	0.9807	0.1505	0.0796	0.1011	0.1432	0.9922	0.0961	0.0508	0.0600	0.0894

**Note:**  $\psi = (\beta, \rho, \lambda, \sigma_v^2)'$ . Parameters values for generating  $x_t$ :  $\theta_x = (.1, .5, .5, 5, 1)$  (see Footnote 13)